AWARD NUMBER: W81XWH-14-1-0042

TITLE: See-What-I-Do: Increasing mentor and trainee sense of co-presence in trauma surgeries with the STAR platform

PRINCIPAL INVESTIGATOR: Juan P Wachs

RECIPIENT: Purdue University,
West Lafayette, IN 47907

REPORT DATE: April 2015

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT:

Approved for public release; distribution is unlimited.

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**
April 2015	Annual	3 Mar 2014 - 2 Mar 2015

4. TITLE AND SUBTITLE
See-What-I-Do: Increasing mentor and trainee sense of co-presence in trauma surgeries with the STAR platform

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-14-1-0042

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Juan P Wachs

Email: jpwachs@purdue.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
.
Purdue University
West Lafayette, IN 47907

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The work this year shows a number of studies which provide a preliminary indication that STAR allows trainees to follow some mentor instructions more accurately. The main finding is that focus shifts were greatly reduced when using STAR as opposed to the conventional system. This is a reasonable result, given that a participant in the conventional condition is required to shift focus in order to access the instruction, while in the STAR condition accessing the instruction does not require shifting focus. Based on the completion of Experiment 1 we can state that on average, the placement error was considerably smaller when using the AR system than when using a separate screen. The tablet provides precise feedback as to where a pointer should be placed and the participant leverages this feedback to minimize placement error.

15. SUBJECT TERMS-
Telementoring, Augmented reality, Training and simulation, Real-time instruction, Trauma surgery

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	81	**19b. TELEPHONE NUMBER** *(include area code)*

TABLE OF CONTENTS

1. **INTRODUCTION:** Narrative that briefly (one paragraph) describes the subject, purpose and scope of the research.

Our primary research objectives are to design, implement, and evaluate a working prototype that enables effective telementoring of a trainee surgeon by a remote mentor. This includes (1) a trainee-site subsystem for augmenting the view of the actual surgical field seamlessly by using a transparent display with illustrations of the current and next steps of the procedure, and (2) a mentor-side patient-size interaction platform with a gesture-based interface.

2. **KEYWORDS:** Provide a brief list of keywords (limit to 20 words).

Augmented reality, telementoring, telemedicine, annotation anchoring, transparent display, surgical training, co-presence, simulation, tele-existence.

3. **ACCOMPLISHMENTS:** The PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency Grants Officer whenever there are significant changes in the project or its direction.

What were the major goals of the project?

List the major goals of the project as stated in the approved SOW. If the application listed milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion.

Specific Aim 1:

Implement transparent display (03-Mar-2014 – 03-Aug-2015) <u>60%</u>
Achieve a visual overlay of info. from the mentor (03-Mar-2015 – 03-Mar-2016) <u>15%</u>
Experimental Design 1: trainee subsystem (03-Apr-2016 – 03-Mar-2017) <u>35%</u>

Specific Aim 2

Develop a gesture-based interaction system (03-Mar-2014 – 03-Aug-2015) <u>50%</u>
Experimental Design 2: Gather gesture set (03-Apr-2015 – 03-Mar-2016) <u>10%</u>
Experimental Design 3: Mentor subsystem (03-Oct-2016 – 03-Mar-2017) <u>5%</u>

What was accomplished under these goals?

For this reporting period describe: 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative); and/or 4) other achievements. Include a discussion of stated goals not met. Description shall include pertinent data and graphs in sufficient detail to explain any significant results achieved. A succinct description of the methodology used shall be provided. As the project progresses to completion, the emphasis in reporting in this section should shift from reporting activities to reporting accomplishments.

Major activities: ***Research, develop, and assess a transparent-display augmented-reality system that allows the seamless enhancement of a trainee surgeon's natural view of the surgical field with annotations and illustrations of the current and next steps of the surgical procedure.***

Specific Objectives

Task 1.1- Implement transparent display

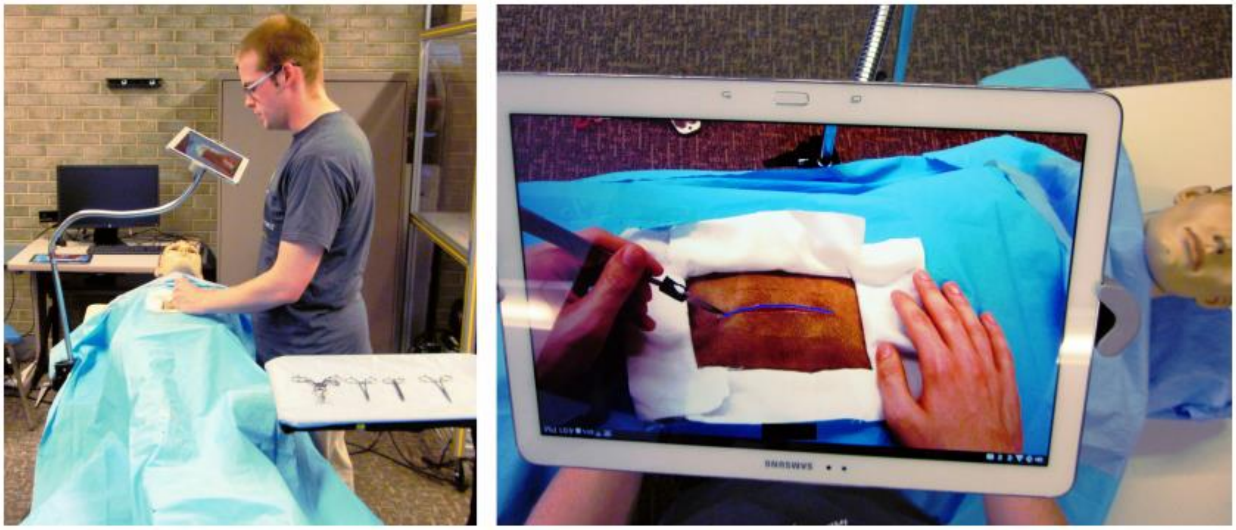


Figure 1: Trainee system in our first implementation of the AR transparent display telementoring approach: overall view (left) and trainee view (right). The trainee surgeon sees the surgical field through the transparent display and performs an incision along the line suggested by the mentor.

Significant results:

- Tablet is suspended in the visual field of the trainee surgeon using a mechanical arm.
- The front facing video camera of the tablet is turned on and the video feed is displayed in real time on the tablet display.
- The setup provides a first implementation of the transparent display: the trainee sees their hands in real time as they operate underneath the tablet (Figure 1).

Conclusions:

- The positioning of the tablet allows the tablet camera to capture a clear view of the operating area.
- By transmitting the recorded video frames, the remote mentor is able to view the operating area clearly.
- A fully simulated transparent display effect is not currently implemented. As the trainee moves his/her head in relation to the trainee tablet, the video frames on screen do not change, meaning that the illusion of transparency can be broken. As shown in Figure 1, there is a mismatch between the portions of the operating field and trainee's hands visible inside the tablet frame, and the portions outside the frame.

Specific Objectives

Task 1.2 - Achieve visual overlay of information

Generation of Visual Overlay – Mentor Tablet User Interface

Visual overlay of relevant surgical information is provided by the mentor to the trainee via a touch-screen user interface on the mentor tablet. The mentor tablet displays the UI on top of the video feed provided by the trainee tablet, so that the mentor can create, delete, and modify annotations and deliver the changes to the trainee. The mentor has access to buttons to create point, line, and loop (closed line) annotations, as well as a toolbox of various sprite-based annotations for surgical instruments, pre-defined text labels, and images of hands in various positions. When placing sprite-based annotations (tools, labels, and hands), the mentor can use multi-touch dragging actions to move, rotate, and scale the tools into place. Any action to edit annotations on the mentor tablet will freeze the screen, to give the mentor a stable canvas to work on.

There is also a button for the mentor to deliver the current annotation state to the trainee system, as well as to clear all existing annotations, or to remove a single selected sprite-based annotation.

We show and describe below an example screenshot of the mentor UI:

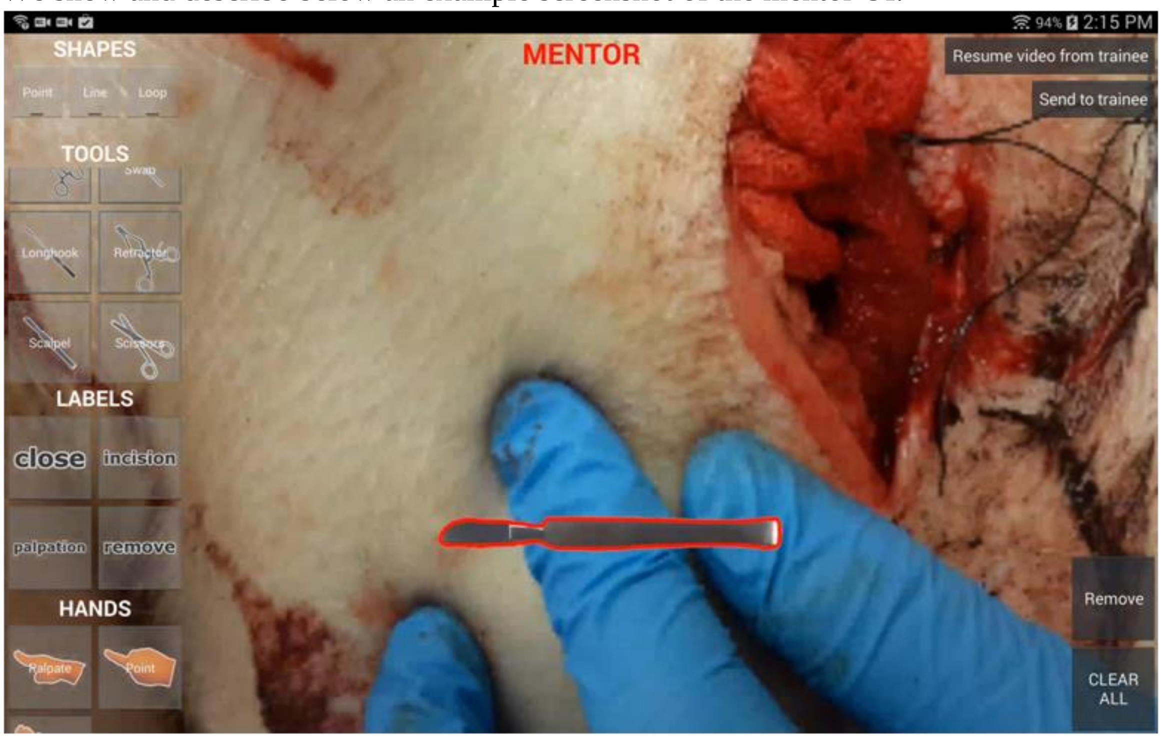


Figure 2: The mentor tablet's interface, showing buttons for drawing shapes, adding tool/label/hand annotations, and sending annotation data to the trainee.

Function name	Description
Point	When selected, allows mentor to create a point annotation by clicking anywhere on the video background.
Line	When selected, allows mentor to create a line annotation by clicking anywhere on the video background, and dragging to draw the line. Lines can be curved depending on how the user moves finger.

Loop	When selected, allows mentor to create a loop annotation by clicking anywhere on the video background, and dragging to draw the loop. A loop annotation is like a line annotation except its ends are connected. Used for drawing borders around objects.
Tool	When selected, allows mentor to create a tool annotation by clicking anywhere on the video background. This kind of annotation appears as an image of a surgical instrument. Can press with one finger to select and drag already-created tool, and can use two fingers to rotate/scale tool in place. Available tool annotations: - BVM (bag valve mask) - Endotracheal tube - Hemostat - Iodine swab - Longhook - Retractor - Scalpel - Scissors - Stethoscope - Surgical tape - Syringe - Tweezers
Label	When selected, allows mentor to create a label annotation by clicking anywhere on the video background. This kind of annotation appears as one of a set of pre-made textual labels. No custom text labels are currently available in the system. Can press with one finger to select and drag already-created label, and can use two fingers to rotate/scale label in place. Available labels: - “close” - “incision” - “palpation” - “remove” - “stitch”
Hand	When selected, allows mentor to create a hand annotation by clicking anywhere on the video background. This kind of annotation appears as a photographic image of a hand in a gesture to represent a certain action. Can press with one finger to select and drag already-created label, and can use two fingers to rotate/scale label in place. Available hand gesture images: - palpate - point - stretch
Resume video from trainee	When the mentor is editing or manipulating annotations on the tablet, the video stream from the trainee pauses to give the mentor a stable working area. The mentor can press this button at any time to resume the live video feed from the trainee.
Send to trainee	When pressed, the system sends data about the currently-created annotations to the trainee tablet, where they will also appear in the same locations.
Remove	Button is enabled only if a tool, label, or hand annotation is currently selected (has a red outline). In this case, pressing Remove will remove that specific annotation.
CLEAR ALL	When pressed, removes all annotations from the screen.

Annotation Anchoring

The first specific aim described in our proposal was to research, develop, and assess a transparent-display augmented-reality system that allows the seamless augmentation of a trainee surgeon's natural view of the surgical field with annotations and illustrations of the current and next steps of the surgical procedure. We have made several improvements to the design and algorithms used in our system, that make its ability to anchor virtual annotations to the surgical field more robust to tablet repositioning and camera occlusion.

Annotation Anchoring Background

To provide informative context for the improvements we have made to our annotation anchoring algorithm, we describe below the general architecture and approach to annotation anchoring.

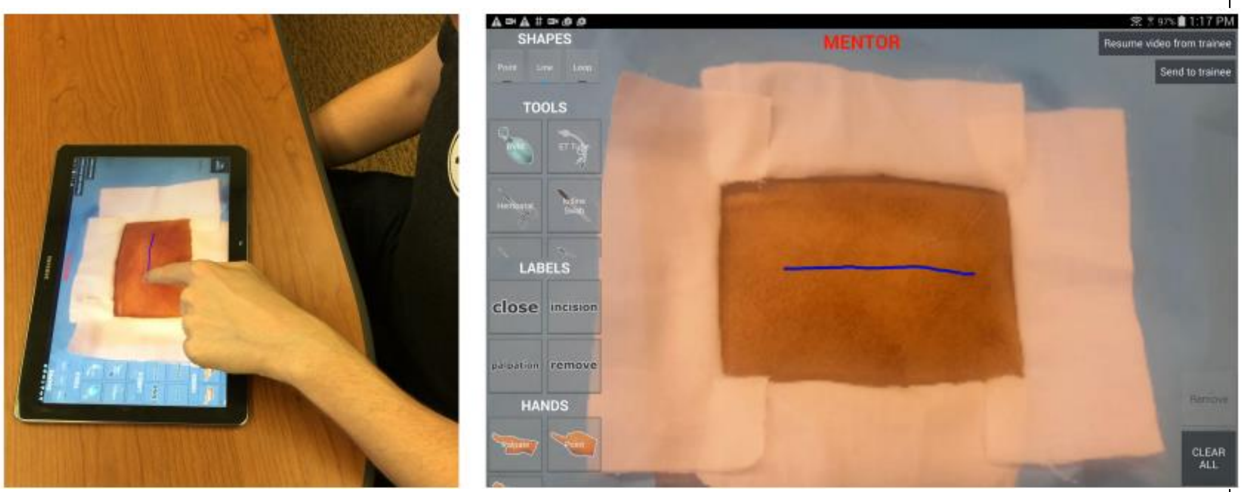


Figure 3: Mentor system: overall view (left) and mentor touch-based user interface (right). The mentor suggests an incision line on the video stream received from trainee system.

Both trainee and mentor interact with their own nearby tablet, as illustrated in Figures 1 and 3. The trainee looks through a tablet suspended between the trainee's head and the operating area, and the trainee tablet captures a video stream of the operating area, delivering it to the mentor. When the mentor tablet receives this video stream on the mentor tablet, the mentor is able to select a frame and augment it with textual or graphical annotations overlaid onto the frame.

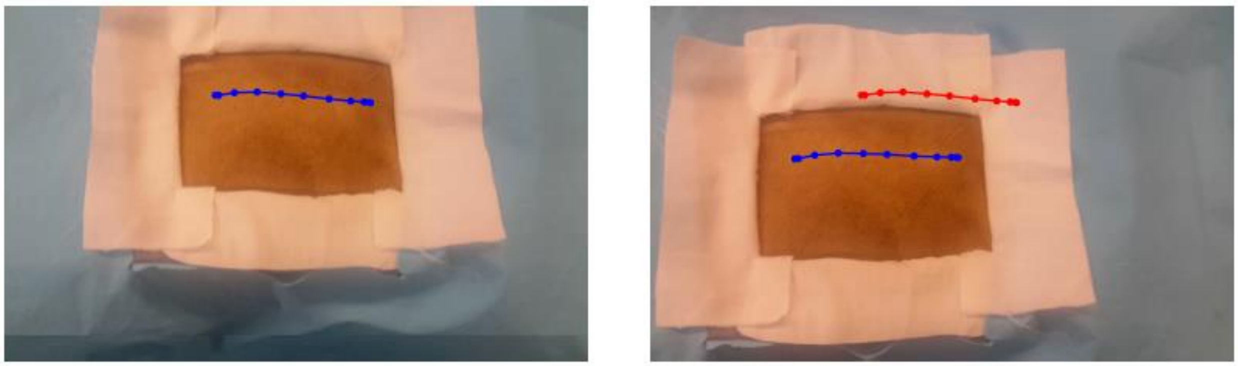


Figure 4: Incision line annotation defined in reference frame with four segments (left, blue), obsolete annotation position in current frame (right, red), and correct position of annotation position in current frame (right, blue).

These annotations will be sent to the trainee tablet to be displayed as a helpful overlay for the trainee user, for surgical guidance. In order for these annotations to remain anchored onto the relevant areas of the physical operating field in subsequent frames, the mentor tablet system must first automatically include information that associates the annotations with nearby salient features of the initial frame (Figure 4). The algorithm to preprocess the reference frame is shown below (Algorithm 1):

input : Reference frame F_0, annotation A defined in F_0
output: ORB features and descriptors of A region in F_0

Compute region R of A in F_0
Detect features f_{0i} in R using ORB
foreach f_{0i} **do**
| compute a descriptor d_{0i} using ORB
end
Return f_{0i} and d_{0i}

Algorithm 1: Annotation anchoring preprocessing of reference frame

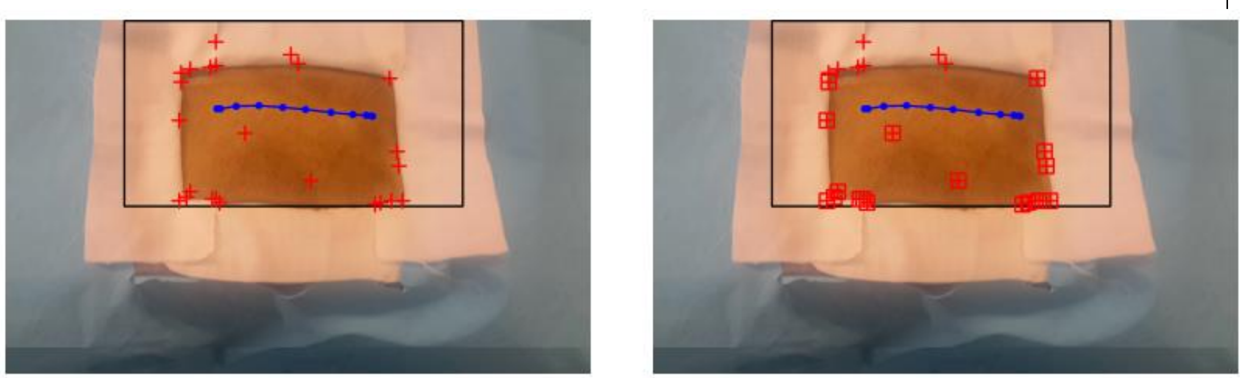


Figure 5: Left: features (red crosses) detected in the reference frame in the region (black rectangle) of the incision line annotation (blue line). Right: descriptors (small red rectangles) computed for features to enable comparison and matching to descriptors in new frames.

This associated information is in the form of features and descriptors (Figure 5). Given the input initial frame, the mentor system uses a computer vision feature detection algorithm to

find salient keypoints around edges and corners in the image. These features are then input into a descriptor extraction algorithm that uniquely identifies the region around each key point. These features and descriptors are sent, along with the annotation's location on screen, to the trainee tablet.

The algorithm for annotation anchoring in the current frame is shown below (Algorithm 2):

input : Annotation A defined in reference frame F_0, ORB features f_{0i} and descriptors d_{0i} of A region in F_0, current frame F
output: Frame F with A overlaid at correct position

Detect features f_j in F using ORB
foreach f_j **do**
  compute a descriptor d_i using ORB
end
foreach d_{0i} **do**
  $d_{0i}.matchIndex = 0$
  $d_{0i}.matchDist = HammingDist(d0i, d0)$
  foreach d_j **do**
    if $d_{0i}.matchDist > HammingDist(d_{0i}, d_j)$ **then**
      $d_{0i}.matchIndex = j$
      $d_{0i}.matchDist = HammingDist(d_{0i}, d_j)$
    end
  end
end
$H = RANSACHomography(d_{0i}, d_j)$
foreach *point* p_i *of* A **do**
  $p'_i = Hp_i$
end
Render A with points p'_i in F
Return F

Algorithm 2: Algorithm for annotation anchoring in current frame

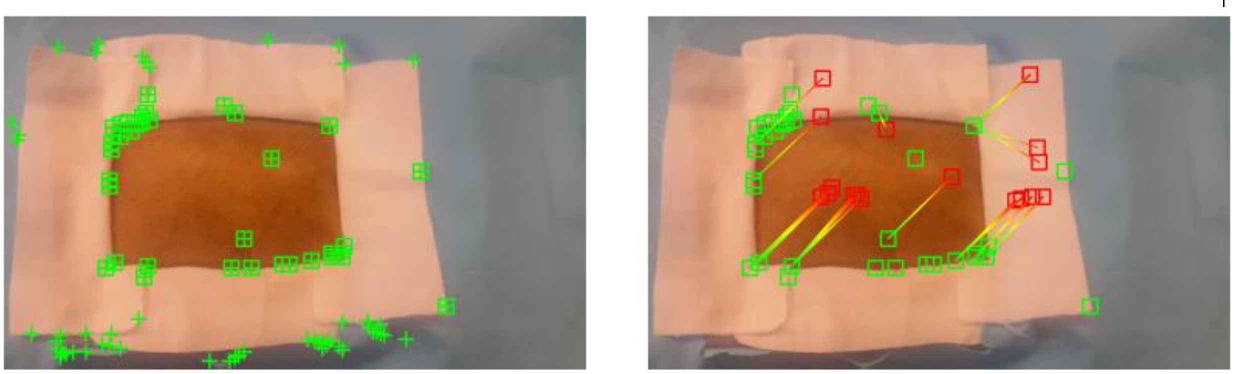


Figure 6: Left: features (crosses) and descriptors (rectangles) in the current frame. Right: reference frame descriptors (red rectangles) matched to current frame descriptors (green rectangles).

When the trainee tablet first receives this annotation data, and for every subsequent frame in which the annotation persists, the trainee tablet repeats this feature detection and descriptor extraction process on the current frame (Figure 6, left). At this point, the trainee tablet has the features/descriptors of the frame in which the annotation was first defined, and the

features/descriptors of the current frame. It then uses a descriptor matching algorithm to find, for each key point in the initial frame, the most similar key point in the current frame (Figure 6, right).

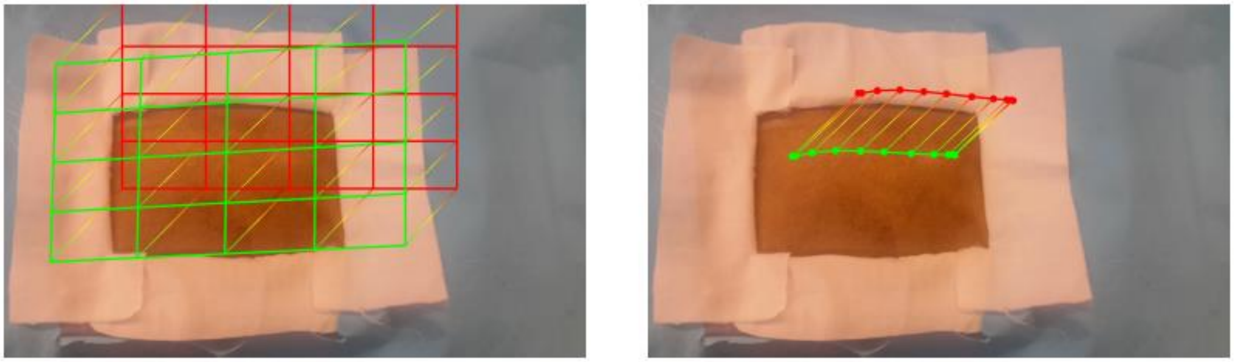


Figure 7: Left: homography linking reference frame to current frame, visualized for a regular grid defined in the reference frame (red) that is mapped to the current frame (green). Right: annotation is anchored by mapping the annotation points from the reference to the current frame.

Once all these matches are found, a RANSAC-based algorithm is used to compute a homography between these two frames -- an affine matrix transformation that describes how to transform points in the initial frame to anchored locations in the current frame (Figure 7, left). One job of this homography-finding algorithm is to eliminate outliers, because some keypoint matches do not correspond with physically associated areas in each frame.

After a homography is found, the trainee tablet applies the homography to each coordinate in the annotation's initial data, generating new coordinates at which to draw the annotation for it to appear correctly anchored (Figure 7, right).

Annotation Anchoring Improvements

Recent algorithmic improvements to our system's annotation anchoring process primarily involve the feature detection stage, and the homography computation stage.

For the feature detection stage, we had initially used the FAST feature detection algorithm to find salient features. While this algorithm returns a large number of features and is computationally efficient, the features are lower-quality in the sense that they are susceptible to noise in the video frames and are less likely to appear in subsequent frames. Furthermore, they are not robust to scaling, meaning that if the tablet is repositioned further away or closer to the operating area, the same features may not be found in subsequent frames. We have revised this step to instead use the ORB algorithm for feature detection. ORB is a variant of FAST that uses image pyramids; this means that it detects features on differently-sized instances of the image, resulting in a set of features that is smaller but also more likely to be robust to scale variation in subsequent frames. ORB is also slightly slower than FAST, but the difference is offset by resulting speedups in descriptor matching: when both initial and current frame have fewer features/descriptors, it is less computationally intensive to find matches between them.

In the homography computation stage, we found anchoring improvements by adaptively scaling the RANSAC reprojection threshold based on our image downsample factor. This reprojection threshold is a parameter to this algorithm that establishes an upper error bound for considering a match an inlier or an outlier. The RANSAC algorithm works by iteratively taking a random sampling of matches, generating a homography using them, then testing the proposed homography against other unused matches. If the initial frame point locations in the matches, when transformed by the homography, end up far away from the current frame point locations in the matches, the algorithm determines that some of the points it used for homography computation were actually outliers.

When the RANSAC reprojection value is low, the homography computation process will take longer but the output homography is more likely to be valid; when the reprojection value is high, a homography computation may end early but result in a homography that was computed using an outlier. Usually, a threshold value between 1 and 10 pixels is considered reasonable. Our initial system used a constant reprojection threshold value in this range. However, as part of our annotation anchoring process, we also downsample the input frame by a factor of 4 for efficiency purposes; a constant reprojection threshold value now is too large for this smaller image. By adjusting our reprojection threshold parameter to scale linearly with our downsample factor, the homography computation process remains acceptably strict, returning homographies that are less likely to be computed using outlier matches.

Significant results:

- Developed annotation anchoring process that is able to run on portable tablet systems.
- Annotations are able to be created, modified and transmitted by mentor using touch-based user interface.

Conclusions:

- Annotation anchoring process is able to find reasonable homographic transformations for annotations in scenes with moderate numbers of salient features.
- Adjustments to RANSAC reprojection thresholds in the annotation anchoring process allow system to avoid using outliers in homography computation.
- Current annotation anchoring implementation assumes rigid planar surface for transforming annotations; this assumption is unlikely to hold in real-world surgical settings.

Experimental Design 1 - Metrics to evaluate the technical specifications of the system

Annotation Anchoring Results

As a result of these changes to our annotation anchoring system, we have been able to reduce the average error, and increase the success rate, of annotation anchoring. We test our system by creating an annotation over either a scene of a flat anatomical poster, or over a scene of a surgical dummy (Figure 8). We then reposition the tablet through translation, rotation, and zoom, and also occlude and/or deform the area in view of the tablet. After measuring the pixel distance between the annotation as it appears on the screen, and a ground truth position of where the annotation should have been drawn, we are able to determine the average error for each condition. In addition, we define the success rate for each condition, meaning the percentage of frames in which the anchoring error remains below a threshold value of 20 screen pixels.

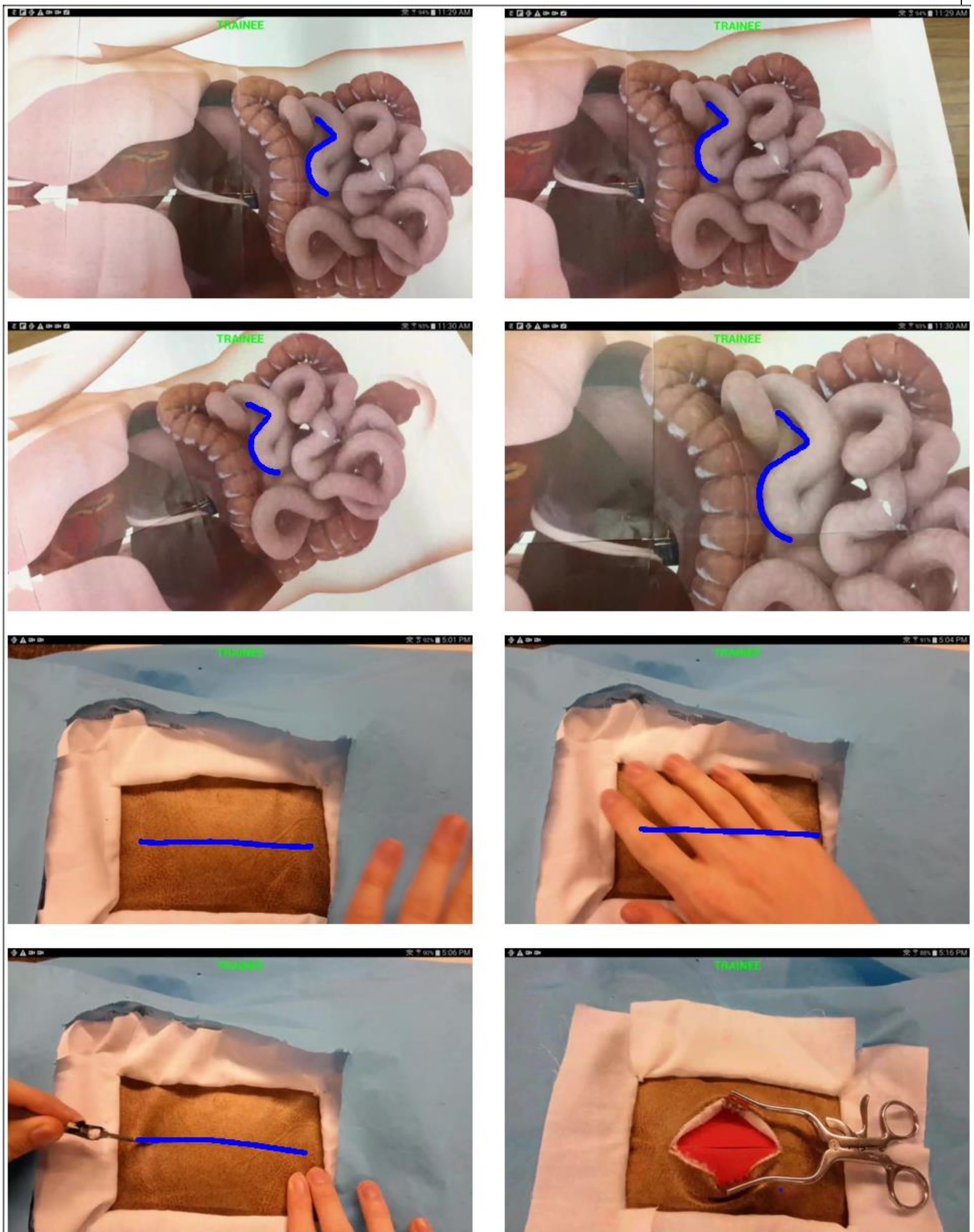
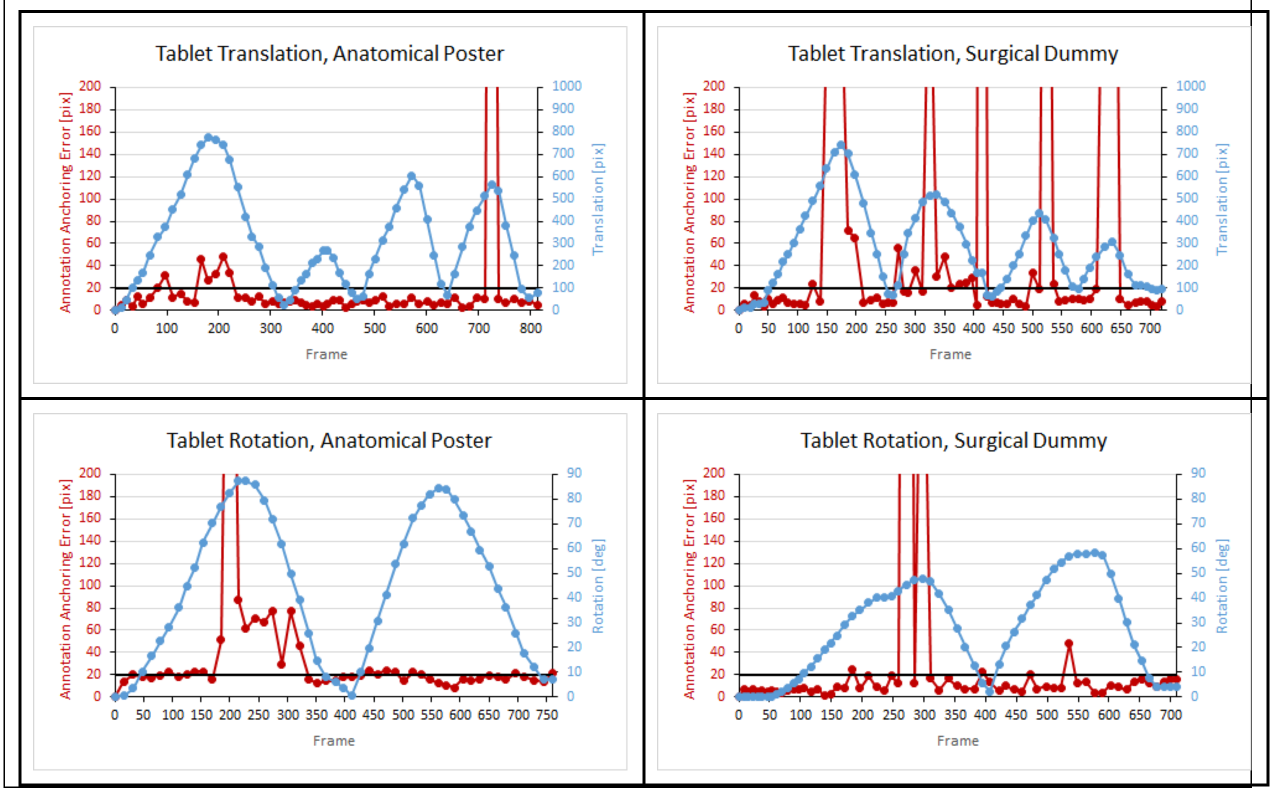


Figure 8: Example frames from our experiments to evaluate the technical performance of our system, both with an anatomical poster (rows 1 and 2), and a surgical dummy (rows 3 and 4).

We show below the current results of our annotation anchoring algorithm, first a chart showing the average error and success rate of each condition (Table 1), followed by graphs showing the error for each frame in our test sequences (Figures 9 and 10). In these graphs, the blue line represents variously the amount of rotation/translation/scaling/occlusion/deformation in the frame, and the red line represents the annotation anchoring error. The black line represents our error threshold.

		Experimental condition						
		Tablet repositioning			Surgical field occlusion		Surgical field deformation	
		Trans	**Rot**	**Zoom**	**Minor**	**Major**	**Small**	**Large**
Scene	**Anatomical poster**	2.66 98%	15.17 55%	7.27 80%	1.7 100%	1.27 60%	n/a	n/a
	Surgical dummy	3.65 89%	8.79 90%	6.41 78%	1.48 96%	3.65 74%	2.90 63%	2.73 15%

Table 1: Average anchoring error in display pixels and annotation anchoring success rate.



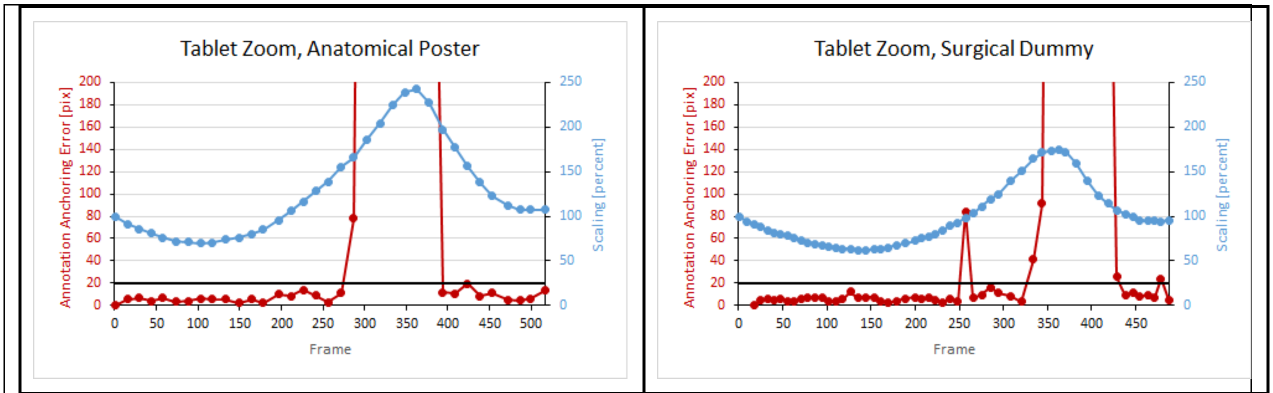


Figure 9: Anchoring error graphs for the tablet repositioning conditions for the sequences from Table 1. The blue lines graph the change in tablet pose, the red lines graph the error values, and the black lines show the error threshold below which tracking was considered successful.

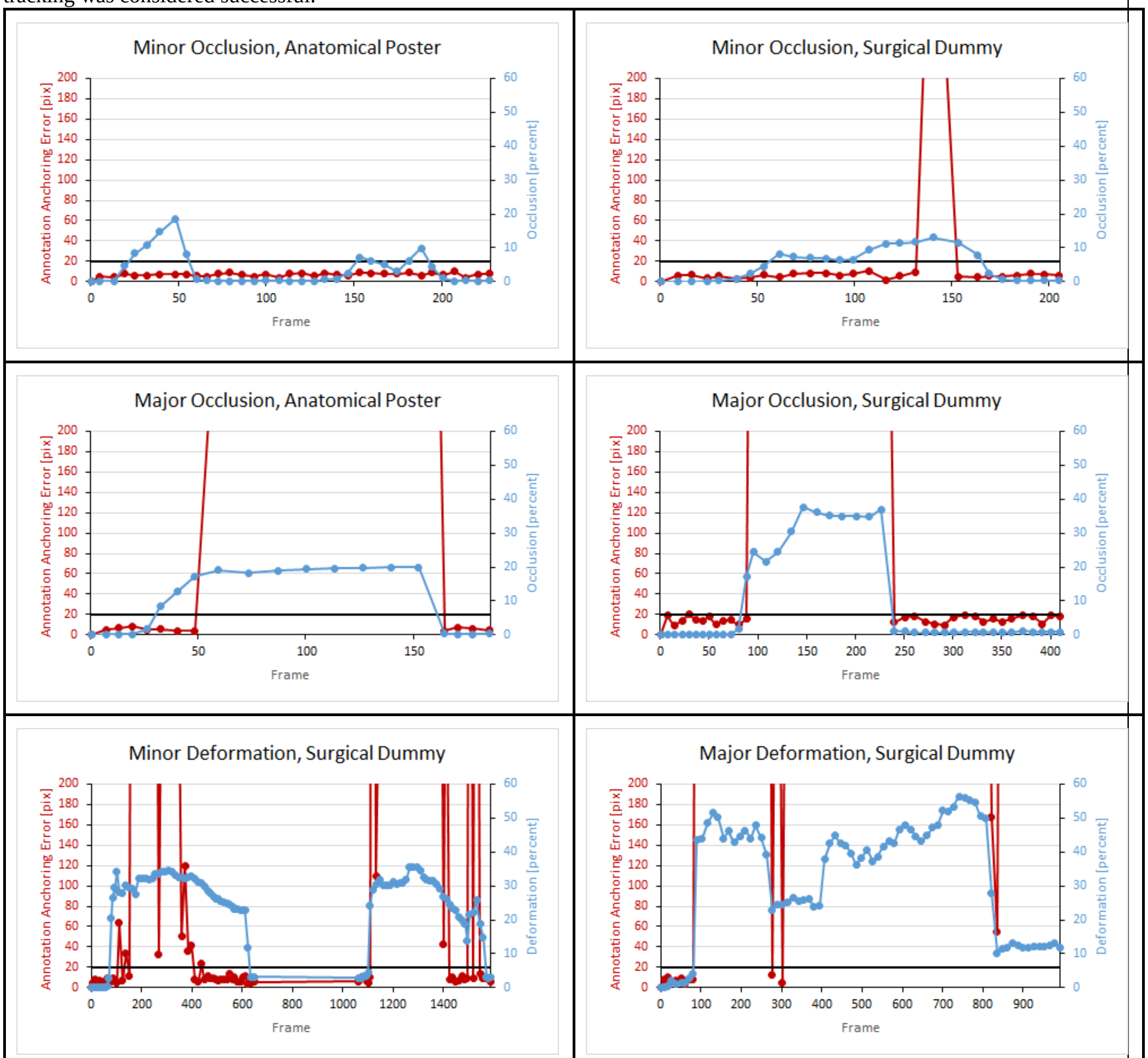


Figure 10: Anchoring error graphs for the occlusion and deformation conditions for the sequences from Table 1. The blue lines graph the amount of occlusion or deformation, the red lines graph the error, and the black lines graph the error threshold.

From these results, we conclude that our system's annotation anchoring algorithms are robust to moderate levels of translation, rotation, and occlusion. Thanks to the scale-invariant features detected by the ORB feature detection algorithm, zooming out does not adversely impact anchoring. However, zooming in can lead to errors, as some of the features in the annotation's initial frame are no longer present to make a robust match. Deformation of the operating environment also continues to be a hard problem for our anchoring algorithm, as it currently only compares the current frame against the initial frame, which can be substantially different in the case of deformation.

System Performance – Speed of Annotation Anchoring

As the STAR system does all annotation anchoring computation on the tablets themselves, without need for offsite computation, the computation time for the anchoring process is very important. The running times for each stage of the annotation anchoring pipeline are recorded below, for various frame resolutions, beginning with full resolution input images, and ending with frames downsampled by a factor of 8 (Table 2).

	Total Frame Time [ms]	**Feature Detection [ms]**	**Descriptor Extraction [ms]**	**Descriptor Matching [ms]**	**Homography Computation [ms]**
1920 x 1080 **1:1**	956	326 34.1%	585 61.2%	2 0.2%	42 4.4%
960 x 540 **1:2**	312	82 26.4%	162 52.1%	2 0.4%	65 20.9%
480 x 270 **1:4**	198	24 12.3%	44 22.1%	1 0.5%	128 65.0%
240 x 135 **1:8**	153	11 7.3%	10 6.6%	1 0.6%	131 85.7%

Table 2: Running times for different stages of the annotation anchoring pipeline, for several input image resolutions.

These results show the importance of downsampling the input frames when doing annotation anchoring. Higher resolution frames imply a larger input to feature detection, and a greater number of features from which to extract descriptors. However, descriptor matching tends to remain a relatively inexpensive operation, and homography computation actually increases in computation time as input resolution gets smaller. This is because, with the RANSAC reprojection threshold getting smaller with input resolution (as described earlier), there are fewer and more error-prone matches, forcing the homography computation to work longer to output homographies that do not use outliers.

Pilot Study

In our proposal, we describe a specific aim (specific aim 3), that we will validate and refine the proposed STAR platform in the context of practice cricothyrotomy procedures on a human-patient simulator in a controlled environment. As part of our research toward validating our platform's ability to benefit trainee users when conducting medical procedures, we conducted a pilot user study (Figure 11). The purpose of this pilot study was to compare the hand-eye coordination, task accuracy and task completion time of participants when using our augmented reality system (the AR condition), compared with using a conventional system for telementoring based on displaying mentor feedback on a nearby monitor (the Conventional condition). We describe the task, conditions, and results below.

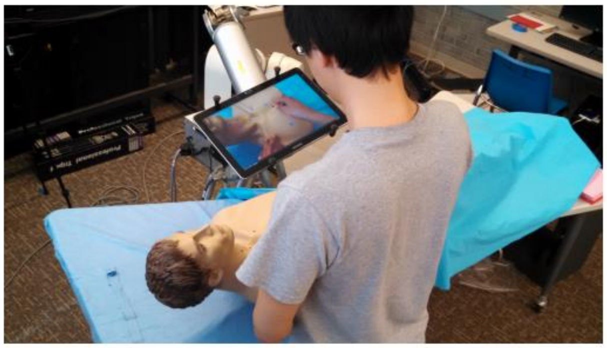


Figure 11: Experimental setup for the AR (left) and Conventional (right) conditions.

Participants

Twenty-two participants were recruited from graduate students of computer science and industrial engineering programs at Purdue University. The participants were randomly divided into two equally-sized groups and assigned to the AR and the Conventional conditions. Each participant wore a Google Glass head-mounted camera, which acquired a video of the task from the participant's point of view.

Task

To simulate testing a trainee's ability to identify regions in the neck area of a patient (a necessary condition for conducting cricothyrotomies), the participants were tasked with placing a set of seven circular paper stickers (each 6.35 mm in diameter) near the neck region of a patient simulator in our lab. The proper locations to place each sticker were provided one at a time by a mentor. This task was repeated for a total of three trials per participant, with each trial varying the location and order of the indicated placement areas. Participants were first given a short (approximately two minutes) verbal description of the task, including a direction to complete the task as quickly and accurately as possible.

AR condition

The participants who used our STAR telementoring system received guidance in the form of virtual annotations appearing on the trainee tablet. The participants were able to look through

the trainee tablet while placing the stickers, receiving live feedback on their hand motions while working. The position of the tablet was kept fixed for all participants using a robotic arm pre-programmed to a particular pose.

Conventional condition

The control condition involved participants receiving sticker placement location instruction from a 46-inch LCD monitor placed near the operating area. These participants would look at the monitor to receive the instruction, and then look back to the operating field in order to complete the requested sticker placement task.

Methods

For all experimental conditions and all participants, we recorded (1) the time each participant took to place all seven stickers, (2) the number and duration of focus shifts, and (3) the sticker placement error. Focus shifts were determined by having each participant wear a Google Glass head-mounted camera while completing the task, and using the recorded camera footage to determine at what points the participant was looking at the operating field, or elsewhere in the room. The sticker placement error was measured in pixels, by overlaying a photograph of the patient simulator after the participant had finished placing the stickers, onto another photograph taken from the same angle that showed the correct reference location of each sticker as communicated to the participants. These photographs were each scaled to a 2560x1600 resolution (the resolution of the tablet).

Results, discussion

For participants in the Conventional condition, the measured sticker placement error was 59.6 pixels on average, with a minimum of 4.3 pixels and a maximum of 467.8 pixels. When using our STAR system, sticker placement error was an average of 32.0 pixels, with a minimum of 1.0 pixels and a maximum of 168.5 pixels. Figure 20 shows a scatterplot of the sticker placement errors, showing that participants in the AR condition tended to be more accurate in their placement. To provide real-world and medically-relevant context to this, this translates to an average error of 0.97 cm for the Conventional condition, and a 0.52 cm average error for the AR condition. According to a surgeon on our team, a reasonable threshold for error when conducting surface-level surgical operations is about 1 cm, which means that participants using our system are able to reduce the number of simulated surgical actions taken that exceed this threshold, bringing the average down to more acceptable ranges.

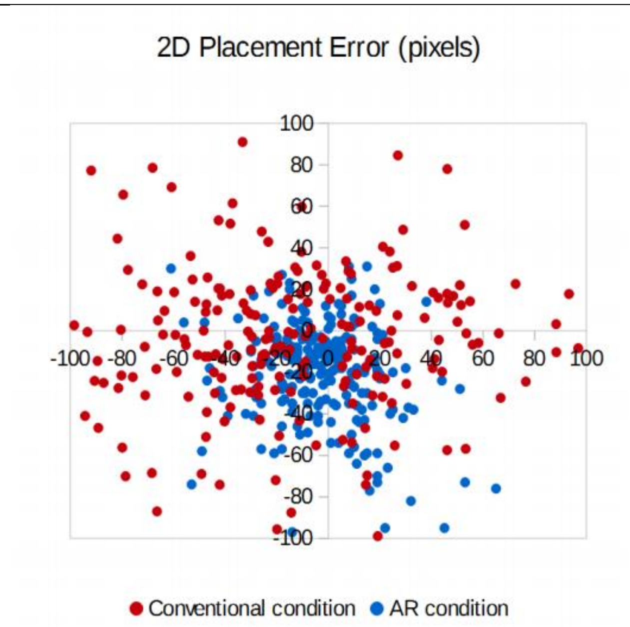


Figure 12: 2D placement error for individual stickers for the AR (blue) and Conventional (red) conditions.

The provides initial support for the third working hypothesis (H3) we described in our proposal, where trainees using our system show higher operative performance compared with using conventional telestrators. An important aspect of operative performance is accuracy of completing a task, which is improved with our system. Our current tests are not in the context of practicing an actual cricothyrotomy, but the sticker placement procedure we have tested in our pilot study is similar because actions were done on the neck area of a patient simulator, and because identification of precise locations in surface-level features is important to both tasks.

Regarding focus shifts when carrying out the experimental task, focus shifts were greatly reduced when using our system as opposed to the conventional system. Participants in the Conventional condition shifted focus away from the operating field an average of 13.8 times, and spent an average of 34% of the operating time looking away from the operating field. In contrast, participants using our STAR system looked away 6.6 times on average, spending an average of 14% of the time with their focus shifted. This is a reasonable result, given that under the Conventional condition, the only way participants could receive instruction was by shifting focus to the nearby telestrator. It should be noted that while some AR participants had 0 focus shifts, most did not; in a few cases participants in the AR condition actually moved their heads to look under the tablet, finding the current lack of a transparent display effect an obstacle to completing the task. Another potential issue here is that the tablet was kept in a fixed position for all AR participants, which may not have been comfortable for varying heights of users. However, in most conditions, the AR participants completed the task with less instances of moving their attention away from the operating area.

This is strong initial confirmation of the first working hypothesis (H1) listed in our proposal: that using this system will result in fewer trainee focus shifts than when using a conventional telestrator. We anticipate that such results will continue to be shown in future user studies involving our system.

An interesting result we found is that task completion time was actually slightly longer for the AR condition than for the Conventional condition. Participants using the Conventional system completed the sticker placement task in 41.31 seconds on average (min=25.7 seconds, max=97.70 seconds), whereas participants using the STAR system completed it in 53.44 seconds on average (min=31.52 seconds, max=80.70 seconds). Some possible causes may be that hand-eye coordination suffers in the AR condition due to lack of depth perception or a fully simulated transparent display effect. However, when we evaluate this in the context of the higher accuracy among AR condition participants, a reasonable conclusion is that participants spent more time when they had immediate feedback and could thus more precisely place the stickers in the correct location. In contrast, participants using the conventional system had no immediate feedback as to whether their proposed placement location was actually accurate, and had no alternative but to go with their first guess.

Significant results:

- Evaluated technical performance of annotation anchoring system, measuring anchoring accuracy against ground truth measurements on planar and non-planar scenes.
- Measured speed of various aspects of anchoring pipeline, and the impact of image downsampling on performance.
- Conducted pilot study that evaluated the STAR system's ability to provide helpful, non-distracting guidance for trainees.

Conclusions:

- Annotation anchoring is robust to tablet repositioning and occlusions. However, it is less robust to cases where the surgical field substantially deforms.
- Pilot study shows that participants using the STAR system performed the requested task more accurately and with fewer focus shifts than participants using a traditional telestrator-based tele-mentoring system.
- Annotation anchoring currently runs at about 12 fps; further improvements may be possible with GPU-accelerated image processing techniques.

<u>Major activities:</u> Research, develop and assess a patient-size interaction platform where the mentor can mark, annotate, and zoom in on anatomic regions using gestures performed over a projected image or on a multipoint-touch screen..

Task 2.1- Develop a gesture-based interaction system

The designed system for gesture-based interaction uses a one-shot learning approach, where the single gesture observation given for training, is used to generate artificial observations through two different methods. The first method the range of motion and the increment between points in a given trajectory, to fit a Mixture of Gaussian (MoG) distribution, used to generate new trajectories; while the second leverages on inverse kinematics and the between-joint angle constraints associated with bio-mechanical constraints of the human body to fit a MoG and generate new artificial trajectories maintaining joint angles around the middle of each range. With the new training set, containing both the original and artificial observations, a Hidden Markov Model (HMM) is trained to detect such gesture in future situations and added to the lexicon set.

Initial interactions with medical experts using the proposed telementoring system, provided some insight to build a lexicon which can effectively relate hand gestures to navigation and

image manipulation actions; commands like zoom, pan, rotate, pick and drop were included in our lexicon. Ten engineering students were recruited to perform 12 gestures, five times each one, for a total of 50 observations per gesture in the lexicon. This dataset was used both to do preliminary studies regarding human arm motion, and helping to determine the static and dynamic joint constraints of a human arm.

Generating Artificial Observations

Method 1: Forward motion propagation using Mixture of Gaussians

Given a gesture trajectory, recorded using Kinect V2 library capabilities for skeleton detection, shown in Figure 1 (left). The concatenation of points in 3D space provides incremental information in all three axis. Considering a vector of dimension d (in this case 3) with N possible increments in a given gesture trajectory $\{X_1, X_2 \ldots X_N\}$, a Gaussian Mixture Distribution Model is fitted using Expectation Maximization (EM) algorithm. The following expression describes the Gaussian Mixture parameters.

$$p(x;\mu_k,\sigma_k,\pi_k) = \sum_{k=1}^{m} \pi_k p_k(x), \quad \pi_k \geq 0, \sum_{k=1}^{m} \pi_k = 1,$$

$$p_k(x) = \varphi(x;\mu_k,\sigma_k) = \frac{1}{(2\pi)^{d/2}\,{\sigma_k}^{1/2}} \exp\{-\frac{1}{2}(x-\mu_k)^T {\sigma_k}^{-1}(x-\mu_k)\}$$

Where m is the number of mixtures in the model, p_k is the normal distribution density with mean μ_k and covariance matrix σ_k which is positive semi definite; π_k is the weight of the k^{th} mixture. Given the number of mixtures, which in this particular case was selected as 3, the algorithm finds the maximum likelihood estimates of all the mixture parameters (μ_k, σ_k, π_k)

$$\Theta = \left\{(\mu_k,\sigma_k,\pi_k): \mu_k \in \mathbb{R}^d, \sigma_k = {\sigma_k}^T > 0, \sigma_k \in \mathbb{R}^{dxd}, \pi_k \geq 0, \sum_{k=1}^{m} \pi_k = 1 \right\}$$

$$\mathcal{L}(x,\theta) = \log(p(x,\theta)) = \sum_{i=1}^{N} log\left(\sum_{k=1}^{m} \pi_k p_k(x)\right) \to \max_{\theta \in \Theta},$$

With the fitted parameters, artificial trajectories are generated from the original one using the expression:

$$artificial_{trajectory} = original_{trajectory} + \sum_{k=1}^{m} \pi_k R_k,$$

Where R_k are random vectors generated using the multivariate normal distributions obtained earlier. Figure 1 (right) shows 10 artificially generated trajectories for each hand in the gesture “Zoom In”.

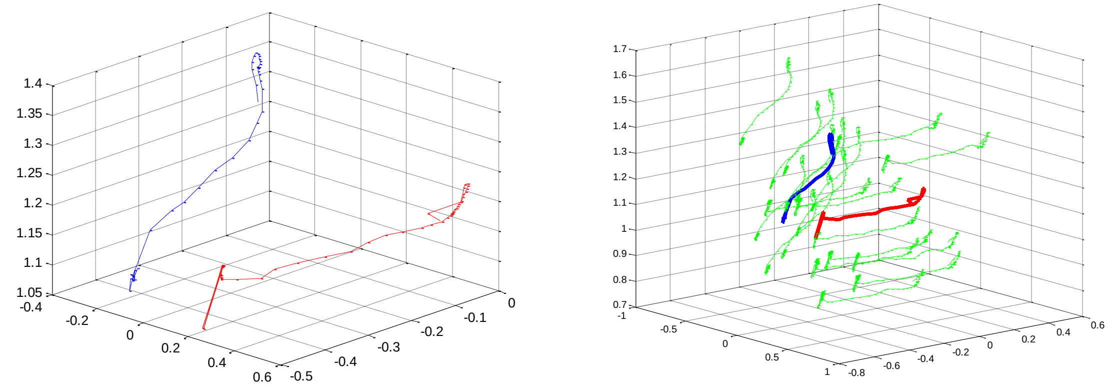


Figure 1 Zoom In Gesture. [Left] Left (Blue) and right (Red) hand trajectories. [Right] 10 artificially generated trajectories for each hand (Green).

<u>Method 2:</u> Backward propagation using bio-mechanical constraints in human arm joints.

The human arm can be modelled as a manipulator with 6 rotational degrees of freedom (DOF) in three major joints (shoulder, elbow and wrist). Each of joints has static and dynamic constraints related to the range of motion. For a healthy person, the static range for each degree of freedom is shown in Table 1.

Table 1 Static Constraints for Range of Motion in Human Arm

Joint	Motion Description	Range (degrees)
Shoulder	Abduction/Adduction *Move arm sideways*	-45 to 180
	Horizontal Extension *Swing arm horizontally forward and backward*	-45 to 130
	Vertical Extension *Raise arm forward and backward*	-60 to 180
Elbow	Flexion/Extension *Move lower arm closer or further away from biceps*	0 to 150
Wrist	Flexion/Extension *Bend wrist closer or away from inner lower arm*	-70 to 85
	Radial/Ulnar deviation *Bend wrist so thumb nears radius or pinky nears ulna*	-20 to 40
	Supination/Pronation *Hand Palm faces up or down*	-90 to 90

Said constraints play an important role when determining the inverse kinematics of the human arm, since some mathematical solutions cannot be reached. Dynamic constraints are also present, when certain values of a given joint limit the range of another. This interaction between joints, presents an opportunity to study the synergies within joints of the arm for a

given motion. If such synergies are modelled, they can be also used to generate artificial trajectories. These new trajectories can incorporate into the system a new level, since the selection of angle joints to actually perform them will keep the angles within the middle of the ranges of motion; this implies a level of comfort and naturalness.

Training HMMs for each gesture

Each HMM is a left-right model, as the one shown in Figure 2, with 5 states. Each HMM is modeled by a combination of matrices, $\lambda_k = (A_k, B_k, \pi_k)$ and trained using trajectories generated through the method previously described. Baum-Welch algorithm was used to tune the parameter matrices on the generated HMMs. A gesture is said to be recognized when a new sequence of observations results with the highest probability among all trained HMM.

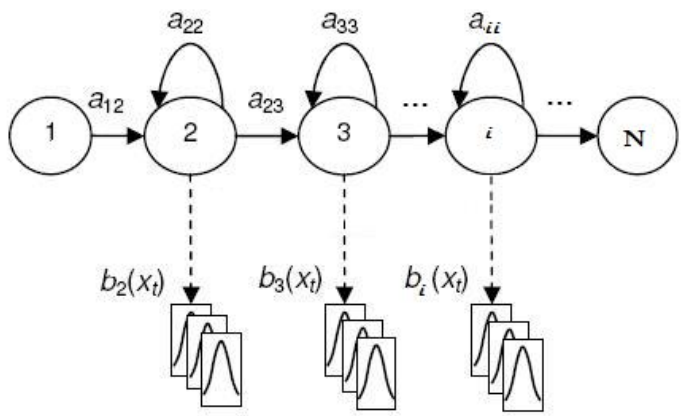


Figure 2 Left-Right model for a HMM with N states

The input information from both hands' trajectories have to be discretized and quantized in the form of observations to tune the parameter B_k in each HMM. Each trajectory is decomposed to become a component of a feature vector. The feature vector is comprised by 3 levels for speed (increasing, decreasing or null), and 18 bins for angle orientation. Each observation will result in a code related to a multi-based combination, shown in the following expression; for a total of possible 1265 observations in each state.

$$obs_{symbol} = [vel_{right} \quad ang_{right} \quad vel_{left} \quad ang_{left}] \begin{bmatrix} 162 \\ 54 \\ 3 \\ 1 \end{bmatrix}$$

Gathering Gesture Set

When developing a gesture-based interaction system, it is important to understand the context of the interaction and the purpose the gestures will convey. For this telementoring project, the context of the gesture interaction relates to the mentor side of the system where they can manipulate a given image received from the trainee's side, and command actions to give a determined instruction. Figure 3 shows some of the gestures included in the initial set, where the mentor may perform basic transformations to an image such as pan or rotate, and can give instructions to pick or drop a given instrument.

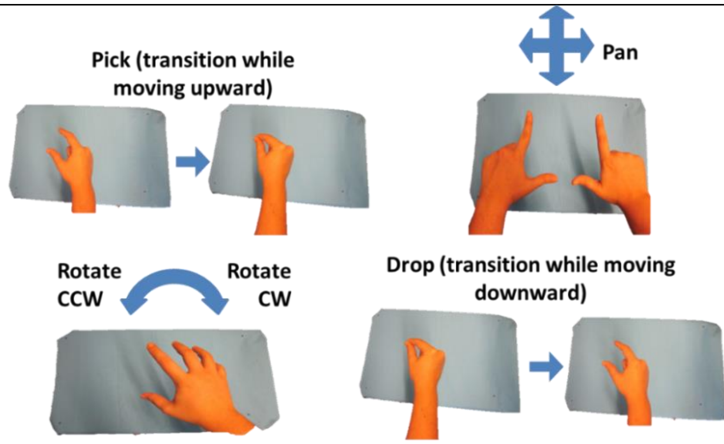


Figure 3 Gestures included in the Initial Lexicon

The method used to capture the mentor's movements includes the capabilities provided by the Kinect V2 libraries, where real-time information of up to 26 joint positions is available. In the context of this project, the skeleton information (shown in Figure 4) limits to the upper limbs since we expect the mentor to be sitting down interacting with a display at table level.

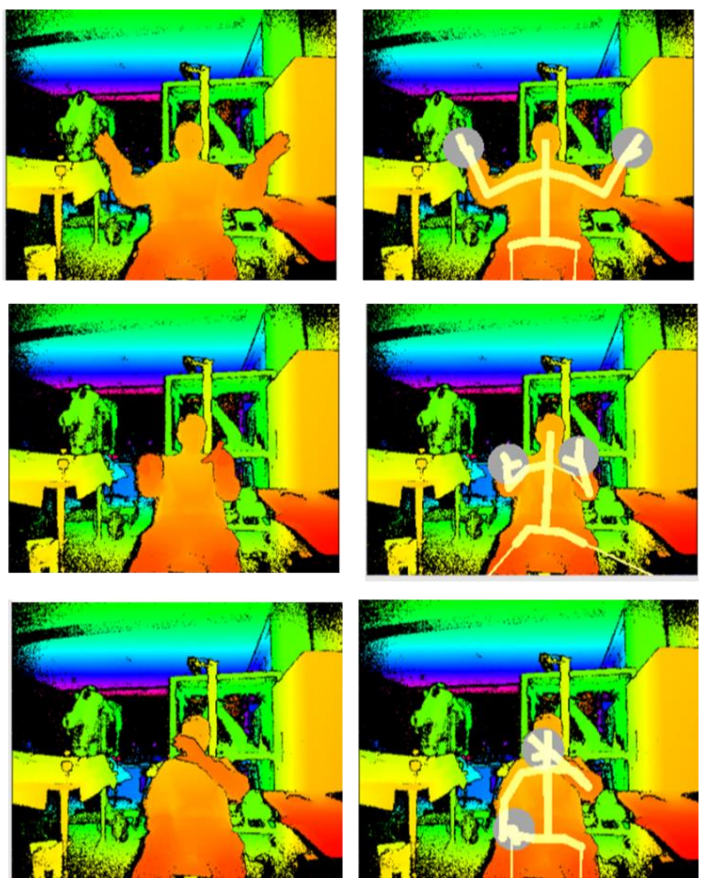


Figure 4 Gestures in Lexicon captured using Kinect V2. [Top] Zoom In. [Middle] Pan Up. [Bottom] Rotate Counter-Clockwise

Significant results:

- Artificial trajectory generation can be used for large or gross gestures by modelling one instance of a trajectory using Mixture of Gaussian Distribution Model.
- Further information needs to be incorporated to the feature vector to account for different hand postures or finger motions.
- Joint synergies may be modelled to generate new artificial trajectories based on bio-mechanical constraints of the human arm.

Task 2.1- The Interaction System

Hardware Design

Our system required a tablet holder, customized to our specifications, as an extension to the WAM™ Arm. We decided to build the tablet holder as an add-on to the Haptics Ball Extension of the WAM™ Arm as it would be relatively easier to build an adapter, like a ball-and-socket joint, onto the ball than any of the other extensions available.

Figure 5: WAM Arm with Haptics Ball Extension

The design criteria determined were:

1. Should be an add-on to the Haptics Ball Extension.
2. Should be easy to attach and detach.
3. Should have at least 3 degrees of freedom to allow the tablet orientation to be flexible.
4. Total cost should not be more than $200.

Preliminary Concept:

The preliminary concept is shown in the image below:

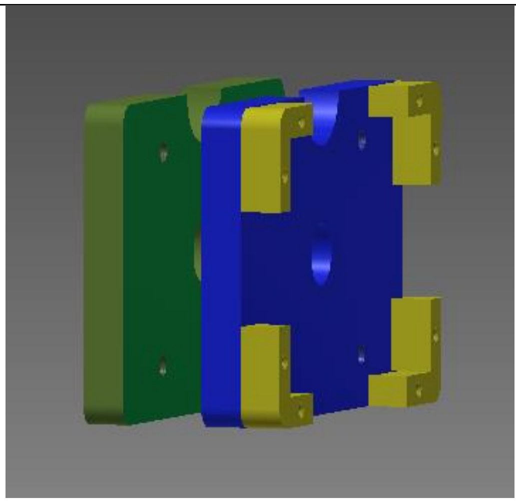


Figure 6: Preliminary Design

It was basically two plates that would "fit" onto the ball and be held in place with four bolts. The image of the top plate below helps explain the concept better:

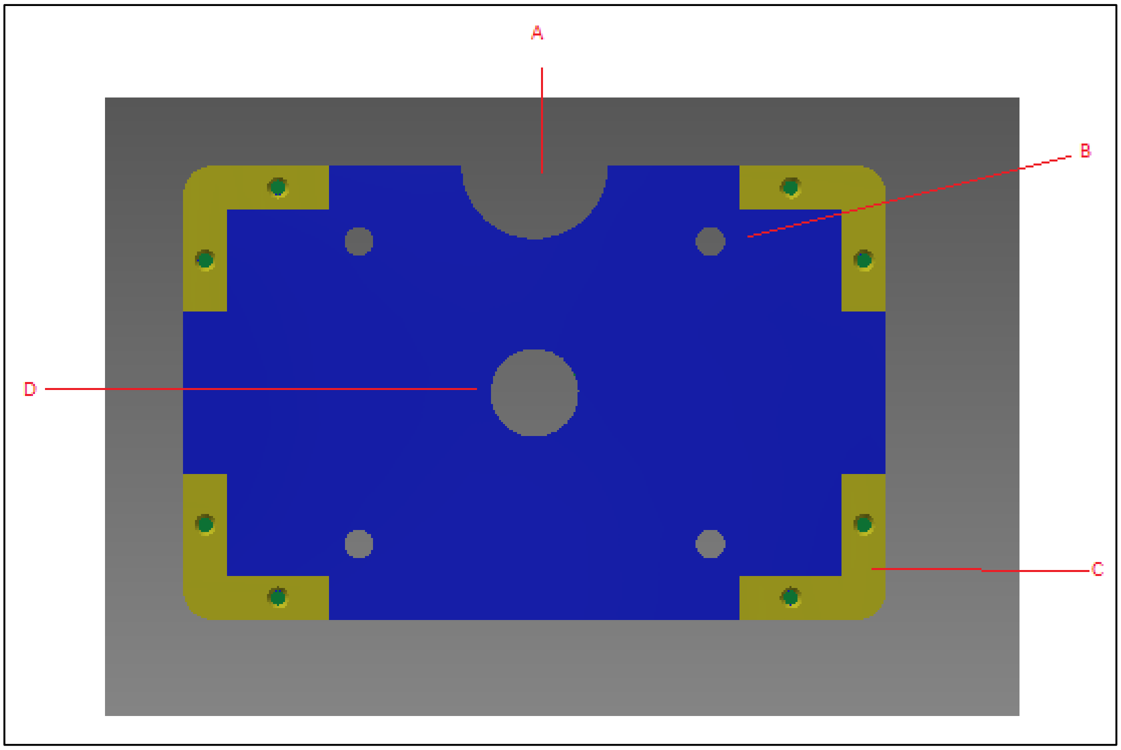


Figure 7: Preliminary Concept (Front View)

A: Crescent shaped hole for the tablet camera
B: Holes for bolts that keep the plates in place
C: Keeps the tablet in place
D: Hole for the plate to "fit" the Haptics Ball

There were many issues with this design. The biggest concern was that the plates would slip on the Haptics Ball.

Final Concept:

Multiple changes to the design and prototypes led to the following final design.
We decided to use the *Atdec Spacedec SDDO Quick Shift Donut Bracket* as the adapter between the Haptics Ball and the tablet holder.

This bracket has a very tight grip. The diameter of the donut can be adjusted via set screws. There is a rubber layer on the inside which adds to the grip and prevents the bracket from slipping on the Haptics Ball.

There is a ball and socket at the base which allows 3-degrees of freedom

Figure 8: Donut Bracket Adapter

We selected the RAM™ X-Grip for holding the tablet itself. The knobs at the four ends are spring loaded and can be used to adjust the grip as needed. The rubber tips provide a very strong grip to hold the tablet. The crescent shaped edges allow enough room for the tablet camera.

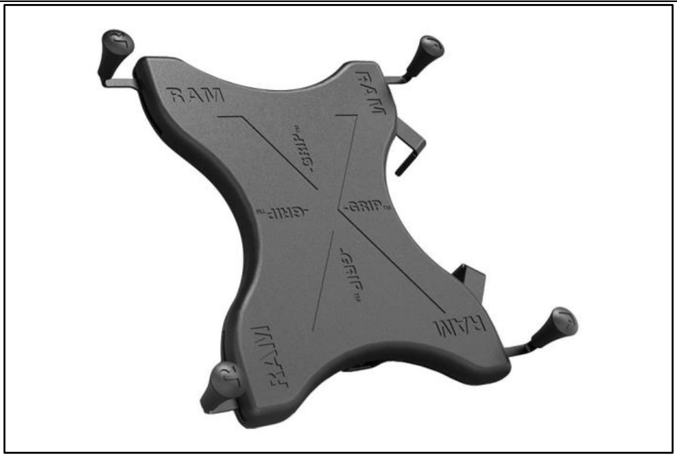


Figure 9: RAM X-Grip

Since both the Donut Bracket and the X-Grip have the same mounting hole pattern it was very easy to assemble them.

Projection Table Design

The mentor table will have a rear projected display. A top-projection will not work as the mentor's hands will occlude light from the projector and interfere with the display.

The following constraints were determined and incorporated in the design for an effective rear-projected mentor display

1. The table-top needs to be made out of a clear material with a polarizing film laminated on the top surface.
2. The table height should be adjustable.
3. There should be little or no frame footprint
4. Should be relatively easy to transport
5. Should be cheap

Table Designs

We chose clear acrylic for the table surface as it is cheap, readily available and very sturdy.

After multiple experiments with two polarizing films; *Vikuiti™ Reflective Display Film* and the *SpyGlass Display Film* from *3M Display*.
We chose the ***Vikuiti™ Reflective Display Film*** since it provided much better resolution.

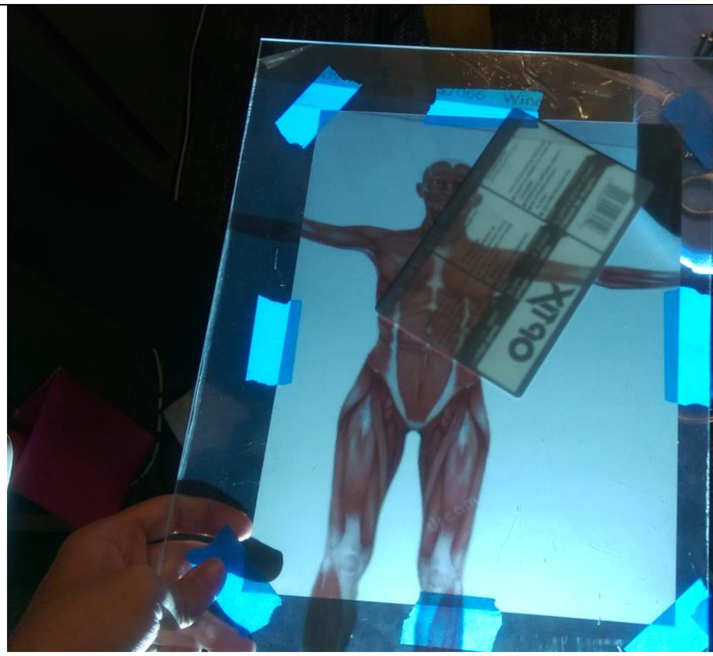


Figure 10: Vikuiti Sample Test Image

Projector

The success of this table depends immensely on the choice of projector. We chose the *BenQ MX842UST XGA Ultra Short Throw Projector* (Shown in the figure below). This projector is relatively cheap, has great resolution and most importantly has a very short throw ratio of 0.47 (If the screen is 3ft away, the projected image will be 6.38ft). The throw ratio is important as the table has to be at a height comfortable for the mentor to work over.

Figure 11: Short Throw Projector

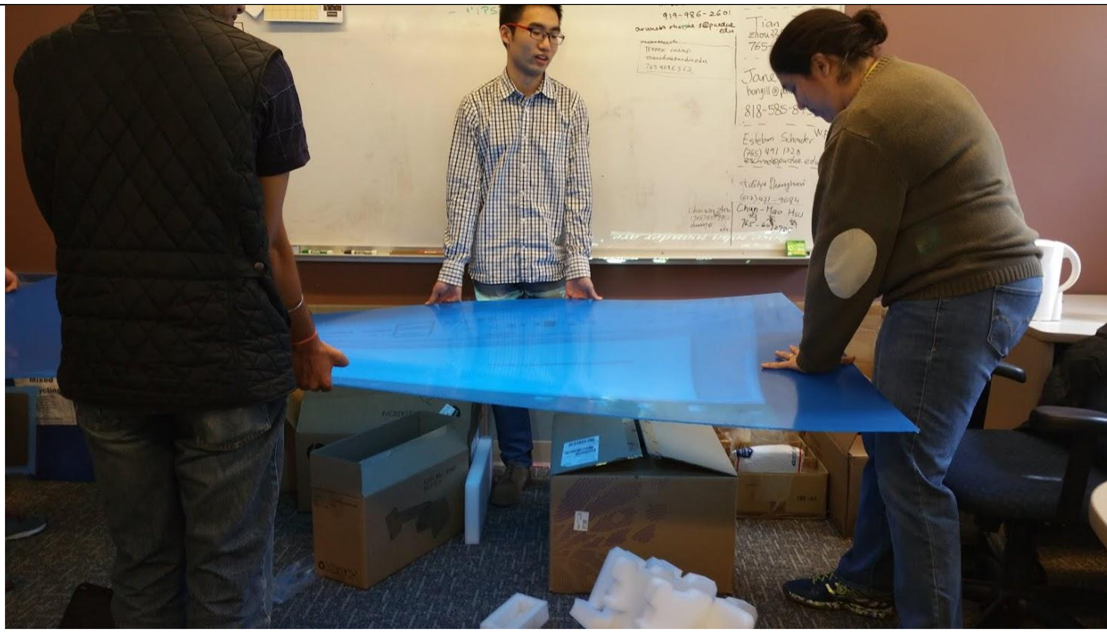


Figure 12. Grad students testing the projection surface

What opportunities for training and professional development has the project provided?
If the project was not intended to provide training and professional development opportunities or there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe opportunities for training and professional development provided to anyone who worked on the project or anyone who was involved in the activities supported by the project. "Training" activities are those in which individuals with advanced professional skills and experience assist others in attaining greater proficiency. Training activities may include, for example, courses or one-on-one work with a mentor. "Professional development" activities result in increased knowledge or skill in one's area of expertise and may include workshops, conferences, seminars, study groups, and individual study. Include participation in conferences, workshops, and seminars not listed under major activities.

Training:

The PI Wachs and co-PI Popescu are mentoring three graduate students: (1) through a research assistantship, computer science PhD student Dan Andersen is working on all the problems related to computer vision, computer graphics and augmented reality under the mentorship of co-PI Popescu; (2) through a research assistantship, industrial engineering PhD student Maria Eugenia Cabrera is working in the problems related to gesture recognition and surface interaction and experimental design, under the mentorship of PI Wachs; (3) through a "guided research" course, ME master student Aditya Shanghavi worked in the development of the fixture to attach the tablet to the surgical bed and attach the tablet to the WAM robot; (4) through a "guided research" course, ME undergrad student Aviran Malik is working on the development of the projection surface for the mentor, together with the help of Aditya Shanghavi; (5) through a "guided research" course, ECE PhD student Chun-Hao Hsu (Chuck)

is developing the software to control the robotic arm WAM to allow free access to the tablet through the robot.

Professional development:

The PI Wachs and the co-PI Popescu participated in the second phase (observed) in three opportunities of the Advanced Trauma Operative Management (ATOM) course at Ashkenazi Hospital (IUSM), with the assistance of Dr. Gerry Gomez, Sherri Marley and Dr. Brian Mullis. The ATOM course demonstrates the surgical repair of common penetrating injuries; 1) a didactic session composed of six standardized 30-minute lectures that review the basic principles of trauma laparotomy and damage control and management of abdominal and thoracic injuries including injuries to the heart and major vessels and; 2) an operative porcine laboratory experience where the human operating room is replicated and surgeons perform operative repairs on 50 kilogram swine. Operative skill is evaluated in the laboratory.

Conferences:

Presented preliminary results of the STAR system:

> An augmented reality approach to surgical telementoring. Loescher, T. ; Shih Yu Lee ; Wachs, J.P. 2014 IEEE International Conference on Systems, Man and Cybernetics (SMC), DOI: 10.1109/SMC.2014.6974276. Publication Year: 2014 , Page(s): 2341 - 2346

How were the results disseminated to communities of interest?

If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe how the results were disseminated to communities of interest. Include any outreach activities that were undertaken to reach members of communities who are not usually aware of these project activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.

The results were disseminated in a talk that the PI Wachs gave in Doha, Qatar, at the Heart Hospital, Hamad Medical Corporation. 11/20/2014. This was presented at the weekly meeting where cases related to surgery are discussed with endovascular surgeons. The focus of the presentation was to discuss the general applicability of this type of technologies to other countries, that are smaller than the US, but with similar needs in terms of surgical care.

On November 11, 2014, another presentation was given at the Purdue CGVLAB (Computer Graphics and Visualization Lab) weekly Graphics Lunch. The audience of the Graphics Lunch presentations are members of the Purdue graphics lab and students and faculty from the university interested in computer graphics and vision. In this presentation, it was described our

vision for the STAR system, as well as details of the current prototype system we have developed. In particular, it was described our approach to annotation anchoring, describing in detail our descriptor matching algorithms for correctly positioning annotations on screen. I was also showed current results in annotation anchoring accuracy, system performance, and formative user feedback from surgeons.

A second presentation at the CGVLAB Graphics Lunch on February 25, 2015. In this presentation a detailed incremental improvements made to the annotation anchoring algorithm was presented, particularly to the feature detection and homography computation stages of the anchoring pipeline. It was provided a survey of the state of the art in simulated transparent display research. It was described several avenues of ongoing research, including using the Google Project Tango tablet's depth camera to acquire geometry of the operating field, and also implementing consensus-based matching and tracking of keypoints to further improve annotation anchoring accuracy.

Presentation of the project given at the annual meeting of the Urological Society of India as part of an named lecture/oration "Present and future of Urologic Robotic Surgery".

Co-PI Popescu described project goals and results to computer science undergraduate class taught in fall 2014. This provided an exposure to research for undergraduate students who are not normally involved in research.

Co-PI Popescu initiated the Augmented Reality Tea (ART) weekly meeting where first and second year computer science graduate students learn about fundamental research challenges and applications of augmented reality. The AR transparent display developed by the project was used as a case study. ART is attracting talented computer science graduate students to AR research.

What do you plan to do during the next reporting period to accomplish the goals?
If this is the final report, state "Nothing to Report."

Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.

Task 1.1 – Implement transparent display

In order to further the goals of implementing a truly simulated transparent display effect, we plan to develop a software framework that allows for incremental improvements in the transparent display effect. First, the current video frames acquired by the trainee tablet will be reprojected to simulate a transparent display from a fixed simulated trainee vantage point, and for objects assumed to be infinitely far away. This will give the impression that, when the user views the tablet from a fixed relative position, and assuming no disparity from very close objects, there is no perceptible mismatch between the visible area displayed on the tablet and the real-world view of areas outside the tablet. Once this is complete, this prototype

framework will be augmented with the ability to manually adjust the position of the simulated trainee vantage point. Next, we will integrate the use of a camera that captures the trainee's face as he/she uses the system. This camera will either be the built-in back-facing camera of the trainee tablet, or an additional peripheral camera. This will use existing face-tracking computer vision algorithms to automatically determine the user's perspective.

We will also begin investigation into the use of depth cameras such as the Google Project Tango tablet, which is able to capture point cloud data. By capturing depth information we plan to develop techniques to incorporate scene geometry into the transparent display effect, which will be vital for simulating transparency in a near-camera non-planar scene like the operating field.

Task 1.2 – Achieve visual overlay of information

We will continue to make improvements to the annotation anchoring algorithms, particularly by incorporating the use of intermediate frame data. Currently, each frame is compared only with the initial frame in which the annotation was first defined. By implementing and experimenting with consensus-based tracking approaches, which combine descriptor matching with continuous methods such as optical flow, we anticipate to be able to improve annotation anchoring.

At the same time, we plan to make performance improvements to speed up computation of annotation anchoring. In particular, we plan to use GPU-accelerated techniques for certain portions of the image processing pipeline like feature detection, which is currently unimplemented in the mobile versions of the computer vision libraries we are using.

Experimental Design 1 - Metrics to evaluate the technical specifications of the system

As before, we will continue to make technical evaluations of the annotation anchoring accuracy and performance of our system, as we make the aforementioned improvements. We will test our anchoring system on the same input image frames as before, to be able to quantitatively determine if our changes improve performance.

In addition, we plan to conduct an extended user study, this time with pre-med and medical students at Purdue University. For this study, we will test these students with the sticker placement task done previously with non-medical students, but participants will also perform a simulated surgical incision task using medical instruments on a patient simulator. We will compare participant performance when using our STAR system for mentee guidance, with performance when using a traditional telestrator-based system. As before, we will measure task completion time, task accuracy, and the number of focus shifts.

4. **IMPACT:** Describe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project relative to:

What was the impact on the development of the principal discipline(s) of the project?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe how findings, results, techniques that were developed or extended, or other products from the project made an impact or are likely to make an impact on the base of knowledge, theory, and research in the principal disciplinary field(s) of the project. Summarize using language that an intelligent lay audience can understand (Scientific American style).

According to our preliminary results, the use of the augmented reality display leads to fewer changes in focus of attention and higher accuracy. These two finding lead us to believe that this technology will increase the sense of co-presence in the operating room between mentor and trainee. This is a fundamental step towards telexistence. Telexistence is a concept used to describe the framework that allows humans to have a real-time sensation of being and interacting with objects in places somewhere different from their actual location. The fundamental premise is that a higher sense of co-presence has an impact on the quality of mentorship. For example, by allowing the mentors to physically interact with the patient's anatomy though hand gestures (embodied interaction), the mentor's level of immersion and engagement will be significantly increased.

What was the impact on other disciplines?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines.

An initial formative experiment conducted during the ATOM surgical training provided some initial understanding about embodied interaction in high risk/ high stakes scenarios. Improved understanding of the factors affecting design and use of embodied interfaces as well as the physical and cognitive requirements for this interaction will be crucial to introduce physical interaction with devices in the OR. We expect a significant breakthrough in this knowledge after our second experimental design, which consists of collecting the gestures that mentors perform while interacting with the large projection table. It is expected that the use of gestural interfaces and the gesture lexicon design will increase the understanding about the different uses of nonverbal communication in the operating room, with extensions to other high-risk/ high-stakes scenarios.

What was the impact on technology transfer?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe ways in which the project made an impact, or is likely to make an impact, on commercial technology or public use, including:

- *transfer of results to entities in government or industry;*
- *instances where the research has led to the initiation of a start-up company; or*
- *adoption of new practices.*

We are currently trying to determine whether this project can result in a patent. For this was have recently contacted the GOR and the PO to see if this is a possible venue to pursue.

What was the impact on society beyond science and technology?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe how results from the project made an impact, or are likely to make an impact, beyond the bounds of science, engineering, and the academic world on areas such as:

- *improving public knowledge, attitudes, skills, and abilities;*
- *changing behavior, practices, decision making, policies (including regulatory policies), or social actions; or*
- *improving social, economic, civic, or environmental conditions.*

Currently the main instrument to improve surgical skills in trauma surgery requires animal models, one to one mentorship and lengthy and complex training sessions (e.g. the ATOM course attended by the PIs of this project). A more cost effective option that will make this training scalable consists of having the training surgeon teach the same ATOM class, remotely, through the STAR platform. This will allow tens residents (current there are only 10-15 per class) to participate concurrently with only one mentor

5. **CHANGES/PROBLEMS:** The Project Director/Principal Investigator (PD/PI) is reminded that the recipient organization is required to obtain prior written approval from the awarding agency Grants Officer whenever there are significant changes in the project or its direction. If not previously reported in writing, provide the following additional information or state, "Nothing to Report," if applicable:

Changes in approach and reasons for change
Describe any changes in approach during the reporting period and reasons for these changes. Remember that significant changes in objectives and scope require prior approval of the agency.

There are no significant changes in the objectives and scope of the project. One simple change is that in Experiment Design 1 instead of using a medical telestrator we are using a large display. This is a good proxy of the telestrator since it allows displaying images and text on top of the images. While the device is different, the functionality and its effect is similar.

Actual or anticipated problems or delays and actions or plans to resolve them
Describe problems or delays encountered during the reporting period and actions or plans to resolve them.

We are expecting to conduct Experiment Design 2 once the projection large table is functional. There was a delay getting the quote for the transparent film (Vikuiti) but now we are in the process of ordering it. We expect to have the table mounted by the beginning of the summer, and conduct experiment 2 during the summer with residents or med students at IUSM.
We also tested Google Glass to use it for displaying the mentor's annotations and found it very inconvenient. We are hoping that Microsoft HoloLens will offer an attractive alternative to Google Glass. Subject to its price we will consider integrating it into our platform.

Changes that had a significant impact on expenditures
Describe changes during the reporting period that may have had a significant impact on expenditures, for example, delays in hiring staff or favorable developments that enable meeting objectives at less cost than anticipated.

No changes.

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents
Describe significant deviations, unexpected outcomes, or changes in approved protocols for the use or care of human subjects, vertebrate animals, biohazards, and/or select agents during the reporting period. If required, were these changes approved by the applicable institution committee (or equivalent) and reported to the agency? Also specify the applicable Institutional Review Board/Institutional Animal Care and Use Committee approval dates.

Significant changes in use or care of human subjects

- Indiana University IRB Approval - 9/26/2014 Protocol #: 1409037680 (Study 2)
- Purdue University IRB Approval – Jun 25, 2014 Protocol # 1403014622 (Study 1 & 3)
- Purdue University IRB Amendment Approval – Jan 8, 2015 Protocol # 1403014622 (Study 1 & 3)
- Indiana University IRB Amendment Approval - 12/29/2014 Protocol #: 1409037680A002 (Study 2)
- ORP and HRPO - HRPO Log Number A-18043.2 - Approval Memorandum – Jan 14, 2015

Significant changes in use or care of vertebrate animals.

No changes

Significant changes in use of biohazards and/or select agents

No changes.

6. **PRODUCTS:** List any products resulting from the project during the reporting period. If there is nothing to report under a particular item, state "Nothing to Report."

- **Publications, conference papers, and presentations**
  Report only the major publication(s) resulting from the work under this award.

  Journal publications. *List peer-reviewed articles or papers appearing in scientific, technical, or professional journals. Identify for each publication: Author(s); title; journal; volume: year; page numbers; status of publication (published; accepted, awaiting publication; submitted, under review; other); acknowledgement of federal support (yes/no).*

D. Andersen, V. Popescu, M. Cabrera, A. Shanghavi, G. Gomez, S. Marley, B. Mullis, J. Wachs. (Major revision). Virtual Annotations of the Surgical Field through an Augmented Reality Transparent Display. *The Visual Computer (under review, first revision);* acknowledged support.

  Books or other non-periodical, one-time publications. *Report any book, monograph, dissertation, abstract, or the like published as or in a separate publication, rather than a periodical or series. Include any significant publication in the proceedings of a one-time conference or in the report of a one-time study, commission, or the like. Identify for each one-time publication: Author(s); title; editor; title of collection, if applicable; bibliographic information; year; type of publication (e.g., book, thesis or dissertation); status of publication (published; accepted, awaiting publication; submitted, under review; other); acknowledgement of federal support (yes/no).*

J. P Wachs. Designing Embodied and Virtual Agents for the Operating Room: Taking a Closer Look at Multimodal Medical Service Robots and Other Cyber-Physical Systems. *Speech and Automata in Healthcare Voice-Controlled Medical and Surgical Robots Series: Speech Technology and Text Mining in Medicine and Healthcare*. A. Neustein (Ed). De Gruyter, 2014; November 2014; ISBN: 978-1-61451-515-9; acknowledged support.

  Other publications, conference papers, and presentations. *Identify any other publications, conference papers and/or presentations not reported above. Specify the status of the publication as noted above. List presentations made during the last year (international, national, local societies, military meetings, etc.). Use an asterisk (*) if presentation produced a manuscript.*

Loescher, T., Shih Yu Lee, Wachs, J.P. An augmented reality approach to surgical telementoring. *2014 IEEE International Conference on Systems, Man and Cybernetics (SMC*), DOI: 10.1109/SMC.2014.6974276. Publication Year: 2014, Page(s): 2341 – 2346. Acknowledged support.

D. Andersen, V. Popescu, M. Cabrera, A. Shanghavi, G. Gomez, S. Marley, B. Mullis, J. Wachs. A Transparent Display for Surgical Telementoring in Austere Environments (2015). Military Health System Research Symposium (MHSRS). (submitted)

- **Website(s) or other Internet site(s)**
  List the URL for any Internet site(s) that disseminates the results of the research activities. A short description of each site should be provided. It is not necessary to include the publications already specified above in this section.

https://engineering.purdue.edu/starproj/
This is the main website of the project, and its main purpose is to disseminate the progress, videos and other visuals and allows the visitors to the website to be exposed to our project. The website is under construction.

https://purr.purdue.edu/projects/starproject/files/
Purdue data repository centralized repository for all data concerned with the project. Data sets, videos, images, results, etc. This website is password protected due to the sensitivity of the data. For access contact the PI Wachs.

- **Technologies or techniques**
  Identify technologies or techniques that resulted from the research activities. In addition to a description of the technologies or techniques, describe how they will be shared

As a result of our research, we have developed and implemented an annotation anchoring technique that is robust to tablet repositioning and to minor occlusion of the operating area. Given a virtual annotation defined in relation to a reference video frame, the algorithm uses feature detection, descriptor extraction/matching, and homography computation in order to determine, for each subsequent frame in the live video, where to reproject the annotation such that it appears physically anchored in the operating field. Details of the algorithm are further described in the section listing current accomplishments, particularly regarding "Task 1.2 - Achieve visual overlay of information". In order to share this technique, we have submitted a journal paper, currently under review, that describes our system and the annotation anchoring technique we use.

- **Inventions, patent applications, and/or licenses**
  Identify inventions, patent applications with date, and/or licenses that have resulted from the research. State whether an application is provisional or non-provisional and indicate the application number. Submission of this information as part of an interim research performance progress report is not a substitute for any other invention reporting required under the terms and conditions of an award.

We intend to fill a patent with the prototype of the STAR system that we developed. This process has not started yet.

- **Other Products**
  Identify any other reportable outcomes that were developed under this project. Reportable outcomes are defined as a research result that is or relates to a product, scientific advance, or research tool that makes a meaningful contribution toward the understanding, prevention, diagnosis, prognosis, treatment, and/or rehabilitation of a disease, injury or condition, or to improve the quality of life. Examples include:
  - *data or databases;*
  - *biospecimen collections;*
  - *audio or video products;*
  - *software;*
  - *models;*
  - *educational aids or curricula;*
  - *instruments or equipment;*
  - *research material (e.g., Germplasm; cell lines, DNA probes, animal models);*
  - *clinical interventions;*
  - *new business creation; and*
  - *other.*

Databases, videos, raw images and recording of the ATOM sessions (3) are located at the PURR repository. https://purr.purdue.edu/projects/starproject/files/

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Provide the following information for: (1) PDs/PIs; and (2) each person who has worked at least one person month per year on the project during the reporting period, regardless of the source of compensation (a person month equals approximately 160 hours of effort). If information is unchanged from a previous submission, provide the name only and indicate "no change."

Name:	*Juan P Wachs*
Project Role:	*Principal Investigator*
Researcher Identifier (e.g. ORCID ID):	*0000-0002-6425-5745*
Nearest person month worked:	*1.12 month*
Contribution to Project:	*Supervising the overall performance of the project. Coordinated visits to IUSM. Working with Maria Eugenia in all the aspects of gesture recognition and one shot learning. Working with Aditya Shanghavi for the design of the large interaction table. Helping with the journal publication.*
Name:	*Voicu Popescu*
Project Role:	*Co-Investigator*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	*1.12 month*
Contribution to Project:	*Actively participated in and advised research assistant Daniel Andersen in the research and development of the first prototype of the augmented reality transparent display surgical telementoring system (i.e. the STAR platform); in designing, conducting, and analyzing the results of user studies aimed at assessing STAR; in disseminating the project results in a journal paper.*
Name:	*Gerry Gomez*
Project Role:	*Co-Investigator*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	*2 weeks*
Contribution to Project:	*Provided formative feedback about the first and second prototype. Conducted the ATOM course and described throughout the course the context of our system. Acted as the mentor in the initial test at IUSM and provided knowledge about the cric procedure.*

Name:	*Brian Mullis*
Project Role:	*Co-Investigator*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	
Contribution to Project:	*Provided formative feedback about the applicability of the prototype to austere environments, and specifically its benefits and drawbacks when used for orthopedic surgery. He also provide assistance regarding the fasciotomy procedure and the possibility to show case this procedure in Experiment 2, in a simulated environment.*
Name:	*Sherry Marley*
Project Role:	*Co-Investigator*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	
Contribution to Project:	*Helped the Purdue team with the experimental design. Coordinated the attendance to the ATOM course three times. She provided consultancy regarding the surgical training process and actionable knowledge during the cric.*
Name:	*Dan Andersen*
Project Role:	*Research Assistant*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	*5.25 months*
Contribution to Project:	*Responsible for architecting, programming and developing tablet system for mentor and trainee tablets. Researched and implemented feature detection / descriptor matching approach for current annotation anchoring algorithm. Was major contributor to journal paper (currently under review) demonstrating the STAR system. Contributed to planning and conducting ongoing user studies to validate system.*
Name:	*Maria Eugenia Cabrera*
Project Role:	*Research Assistant*
Researcher Identifier (e.g. ORCID ID):	

Nearest person month worked:	*5.25 months*
Contribution to Project:	*Maria Eugenia worked together with Dan in the experimental design, recruitment of human subjects, development of the testing environment and mock surgical scenarios. She is now working on the one-shot learning concept for gesture recognition.*
Name:	*Aditya Ajay Shanghavi*
Project Role:	*Master Student*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	*3 months*
Contribution to Project:	*Aditya designed the projection table, tested different projection materials, and types of projectors in order to project a whole silhouette in the table. Aditya also implemented the Gooseneck and the tablet holder and the adaptor to the WAM robotic arm.*
Name:	*Chun-Hao Hsu*
Project Role:	*PhD Student*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	*3 months*
Contribution to Project:	*Chun-Hao programmed the WAM robot gravity compensation feature so it can hold the tablet fixed in one place, but at the same time can take it away when pushed by the hand, or bring back when pulled. Recently he also added voice control.*

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

If there is nothing significant to report during this reporting period, state "Nothing to Report."

If the active support has changed for the PD/PI(s) or senior/key personnel, then describe what the change has been. Changes may occur, for example, if a previously active grant has closed and/or if a previously pending grant is now active. Annotate this information so it is clear what has changed from the previous submission. Submission of other support information is not necessary for pending changes or for changes in the level of effort for active support reported previously. The awarding agency may require prior written approval if a change in active other support significantly impacts the effort on the project that is the subject of the project report.

Juan Wachs	09/01/2014 - 08/31/2017	0.23 SU 0.5 AY
University Of Denver	$200,000	

NSF: MRI Development: Human Avatars: Enabling Research in Natural Communication with Virtual Tutors, Therapists, and Robotic Companions

Major Goals of the Project: The goal of the proposed MRI development project is to develop a life-like emotive software/hardware instrument in the form of robotic character heads that will support natural spoken dialogs between the robot and a human that closely models the face-to-face communication behaviors of a sensitive and effective human tutor, clinician or caregiver to a degree unachievable with current instrumentation.

Overlap: No overlap.

Juan Wachs	09/5/2014 - 08/31/2019	0 SU 0 AY
	$65,000	

NSF: Collaborative Research: I/UCRC for Robots and Sensors for the Human Wellbeing

Major Goals of the Project: The goal of the proposed center is to develop technology in the form of robots and sensors for assistive technologies to support therapies and rehabilitation of people with disabilities.

Overlap: No overlap.

Juan Wachs	04/1/2015 - 03/31/2016	0.12 SU 0.5 AY
	$90,000	

THE NAVSUP FLEET LOGISTICS CENTER SAN DIEGO: An Efficient Real-Time Method for Detection and Characterization of UAVs

Major Goals of the Project: The research objective of this proposal is to develop a video-based methods for real-time detection of small, unmanned aerial vehicles (UAVs) leveraging on effective sense and avoid techniques. Such methods can be integrated into real-time on board processors. This, in turn, would lead to enhanced UAV's capabilities for detection of friendly and unfriendly airborne traffic and respond with appropriate alarms, maneuvers and notifications.

Overlap: No overlap.

What other organizations were involved as partners?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe partner organizations – academic institutions, other nonprofits, industrial or commercial firms, state or local governments, schools or school systems, or other organizations (foreign or domestic) – that were involved with the project. Partner organizations may have provided financial or in-kind support, supplied facilities or equipment, collaborated in the research, exchanged personnel, or otherwise contributed.
Provide the following information for each partnership:
Organization Name:
Location of Organization: (if foreign location list country)
Partner's contribution to the project (identify one or more)

- *Financial support;*
- *In-kind support (e.g., partner makes software, computers, equipment, etc., available to project staff);*
- *Facilities (e.g., project staff use the partner's facilities for project activities);*
- *Collaboration (e.g., partner's staff work with project staff on the project);*
- *Personnel exchanges (e.g., project staff and/or partner's staff use each other's facilities, work at each other's site); and*
- *Other.*

Nothing to Report.

8. SPECIAL REPORTING REQUIREMENTS

COLLABORATIVE AWARDS: For collaborative awards, independent reports are required from BOTH the Initiating PI and the Collaborating/Partnering PI. A duplicative report is acceptable; however, tasks shall be clearly marked with the responsible PI and research site. A report shall be submitted to **https://ers.amedd.army.mil** for each unique award.

QUAD CHARTS: If applicable, the Quad Chart (available on https://www.usamraa.army.mil) should be updated and submitted with attachments.

9. **APPENDICES:** Attach all appendices that contain information that supplements, clarifies or supports the text. Examples include original copies of journal articles, reprints of manuscripts and abstracts, a curriculum vitae, patent applications, study questionnaires, and surveys, etc.

Juan P. Wachs

4 Designing embodied and virtual agents for the operating room: taking a closer look at multimodal medical-service robots and other cyber-physical systems

Abstract: Mistakes in the delivery of health care contribute significantly to patient mortality and morbidity, with an estimate of about 100,000 such cases per year. Some of these mistakes can be directly traced to a lack of effective communication among the surgical team. Studies of verbal and non-verbal communication in the operating theater found that miscommunications frequently occur. While there are other factors that lead to negative case outcomes, such as “team instability” in which teams of nurses and surgeons are not cohesive, or lack of minimal personnel, this chapter will focus specifically on those problems related to lack of communication. This problem is partially solved by the adoption of intelligent sensors along with automation and intuitive technologies in the operating room (OR) to assist surgical teams and improve patient safety. Three different kinds of cyber-physical agents are presented in this chapter. They consist of the Gestix and Gestonurse systems, which are used respectively to assist the main surgeon by displaying patient medical images and in the delivery of surgical instruments, and a telementoring agent that is used during the performance of surgical procedures so as to provide expert guidance to a surgeon in rural areas or in the battlefield.

4.1 Introduction

Mistakes in the delivery of health care contribute significantly to patient mortality and morbidity, with an estimate of about 100,000 such cases per year. Some of these mistakes can be directly traced to a lack of effective communication among the surgical team. In fact, many research studies have found that miscommunications are often the cause of a tragic outcome (Kohn, Corrigan & Donaldson 1999; Firth-Cozens 2004; Lingard et al. 2004; Mitchell & Flin 2008; McCulloch et al 2009; Halverson et al. 2010). Studies of verbal and non-verbal communication in the operating theater found that miscommunications frequently occur. In particular, Lingard et al. (2004) found that requests made in the operating room

are often met with either delayed or incomplete responses, and some of those communications have been found to be directly linked to mistakes in patient care. Of those communications linked to mistakes, the authors found that one third of such communications had a proven detrimental effect on patient health and safety. Halverson et al. (2010) have shown that 36% of these miscommunications were associated with equipment misuse, such as instrument count discrepancies. Egorova et al. (2008) found a strong correlation between instrument count discrepancies and the likelihood that surgical supplies, such as sponges, would be retained in the patient's body.

While there are other factors that lead to negative case outcomes, such as "team instability" in which teams of nurses and surgeons are not cohesive (Carthey et al. 2003), or lack of minimal personnel, this chapter will focus specifically on those problems related to lack of communication. This problem is partially solved by the adoption of intelligent sensors along with automation and intuitive technologies in the operating room (OR) to assist surgical teams and improve patient safety. Three different kinds of cyber-physical agents are presented in this chapter. They consist of the Gestix and Gestonurse systems, which are used to assist the main surgeon by displaying patient medical images and in the delivery of surgical instruments, and a telementoring agent that is used during the performance of surgical procedures so as to provide expert guidance to a surgeon in rural areas or in the battlefield.

4.2 Background

Cao and Taylor (2004) examined how the introduction of robots in the OR to support the surgical team through a surgical procedure presents one way of reducing the number of miscommunications that commonly occur. As simple as it might seem to add automata to the OR to reduce the number of communication problems, there are, however, a number of current roadblocks to the inclusion of robots as teammates in the surgical setting. First, communications among the members of the surgical staff are undoubtedly complex: as such, they involve both verbal and non-verbal expressions (Halverson et al. 2010). How does a robot stand in for a human in a setting punctuated by such complex interactions? In fact, though current speech recognition methods, such as those used in smartphones and tablets, can achieve relatively high recognition accuracy rates, there are still no technologies/algorithms that can deliver a comparable performance when using gaze, gestures and body interaction. Second, robots would need to have comparable performance to existing surgical nurses in their ability to predict

the needs of the surgeon, such as their request for a surgical instrument.[1] Third, since physical interaction is more ambiguous than spoken commands, there is a likely concern that the robot would not be able to distinguish the context in which the physical expression (e.g., gesture) makes most sense. For example, the fist with thumb extended may indicate a request to move the patient upwards, but it can also signify the "OK" sign.

While such communicative challenges, as described above, must be taken seriously, they are by no means insuperable. It is generally agreed that having robotic systems that can address these sets of challenges will enable significant improvements in the OR. An example of such improvements would be for a robotic scrub nurse to be able to recognize the lead surgeon's spoken and nonverbal commands reliably and to be able to promptly identify and fetch the required instrument for the surgeon. The potential miscommunications common to non robotic-assisted surgical teams, would be drastically reduced by placing a robot in the OR who can understand the voice and gesture commands of the surgical team. Furthermore, such a robot would predict with precision the next surgical instrument desired by the surgeon, which would thereby avoid any ambiguous or digressive chains of verbal communications in the OR. Some major benefits might be the shortening of the procedural time for the surgical procedure as well as the cognitive load for the surgeon and his team. In addition, by adding monitoring and wireless communication capabilities to this agent, one can help to reduce the number of retained surgical instruments within the body of the patient. This reduction will be the result of precise monitoring and documentation of instruments used as part of the information stored in the patient's electronic health record (EHR). This has a serious impact on patient safety since retained instruments can puncture internal organs and cause internal bleeding.

Whereas there are those who maintain the point of view that robots as are meant to "replace" jobs, the author suggests the inclusion of robots as helpful collaborators in order to assist the surgeon. Some of the benefits of robotic assistance is the minimization of human errors that are commonly associated with the performance of repetitive and monotonic tasks and the reduction of overall costs. This can be done by incorporating a set of new versatile functions for the robots. Such surgical assistants work in the OR in tandem with the main surgeon, which has been referred to in the literature as a "co-robot." This type of robot is used to cooperate/complement, rather than supplant, the surgeon (Taylor & Stoianovici 2003).

1 Experienced scrub nurses are also known as "mind readers" (Li et al. 2012).

All in all, the introduction of robots augurs well for health care and, more specifically, the OR. There are several ways that this can be demonstrated. First, by improving communication exchanges between the surgeon and his surgical staff, morbidity and mortality can be reduced. Second, it will allow the surgical treatment of conditions that would otherwise not have been affordable. Third, it will lead to a reduction in the actual time spent in the operative and post-operative phase of patient care, thereby reducing costs. Though the adoption of robotic assistants in the OR is still rather new (and study results have shown their general use value in the surgical environment), the author anticipates that in the next couple of years there will be a number of quantitative studies of how such robots may have a positive effect on patient care. Those studies will, thus, prove that the use of robots in the OR significantly reduces mortality/morbidity rates, increases access to surgical care, and lessens time spent in the OR and recovery.

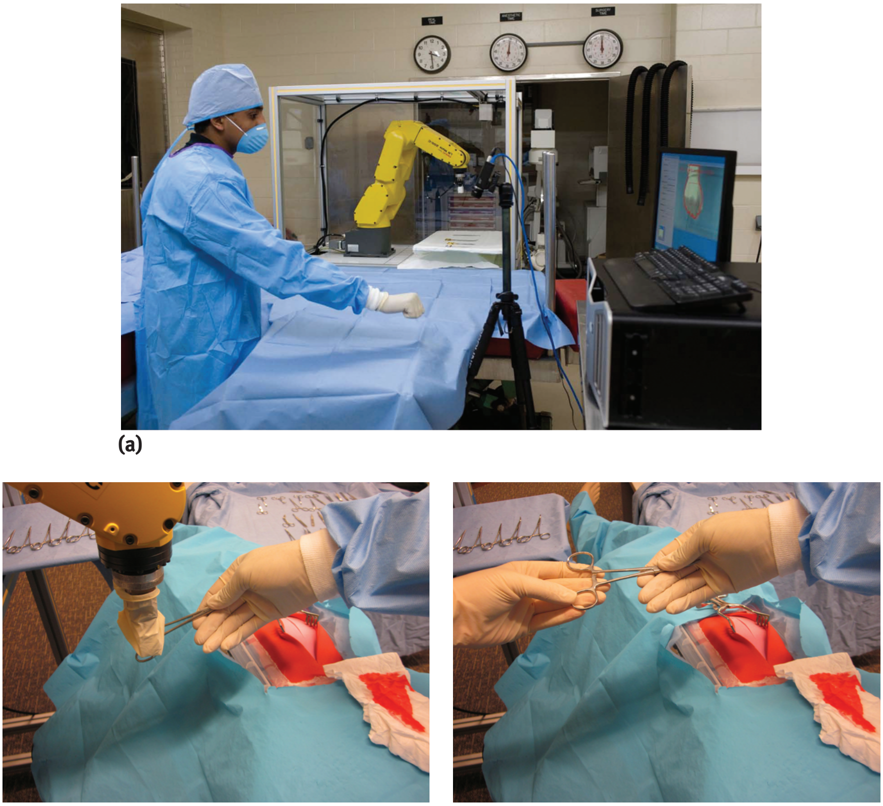
4.3 Design of surgical robots

4.3.1 Types of surgical robots

To better understand the specific role that surgical robots can fulfill, an important distinction must be made between two types of robot assistants. The first type is called *surgeon extenders*. These robots are controlled directly by the surgeon/assistant and they are mainly used to enhance the existing capabilities of surgical instruments and their usability (e.g., certain type of scalpels where the effect of a surgeon's hand tremors is cancelled). The second type of surgical robot is called the *auxiliary surgical support robot*, whose main role is to work side-by-side with the surgical team and assist it in a variety of tasks, such as holding the retractor, or navigating and manipulating the laparoscope tool tip. The later type of robot is often controlled through standard input methods such as pedals, joysticks, speech, and keyboards. The focus of this section is on the second category of robotic assistants – that is, the auxiliary surgical support robots.

Regardless of the type of robot selected, the vast majority of them lack a fundamental recognition of physical forms of expression exhibited in humans and associated with communication events. For example, *Gestonurse* was found to be the only robot that relies on hand gestures combined with voice in order to assist a surgical team during procedures.

Robots that can understand and can interact using nonverbal forms of communication (in addition to verbal forms) can allow the surgeon to interact naturally with the robot without imposing complicated forms of controls. Even more



(a)

(b) (c)

Fig. 4.1a–c: An example of an auxiliary robotic assistant called *Gestonurse* that assists the surgical team: (1a) Interface of the robotic scrub nurse in the OR where the robotic scrub is controlled by the surgeon’s hand gestures (1b) A sterile robot delivering scissors to a surgeon (1c) a surgical nurse (rather than the robot) delivering scissors to the surgeon.

so, these robots would not require their human operators to be re-trained with a new set of commands or controls. Imagine a robot that responds to gestures, body movements, proxemics (the way that humans use the space around them), and speech in a similar manner that surgical nurses do. This type of interaction is a natural and fundamentally sound alternative to traditional forms of interaction, with the advantage that this does not interfere with the normal flow of surgery since their operators would communicate with these agents as if they were interacting with other humans. Though published studies show that robots have been incorporated into the OR as assistants, there is no indication that *non-verbal* interaction between surgical team members and robots constitutes the

main channel of interaction, except for very few exceptions (Webster & Cao 2006; Cunningham et al. 2013).

4.3.2 Challenges and solutions

Successful implementation of automata and intelligent collaboration with such embodied agents involve both technological and societal challenges. Meeting such challenges involve the development of capabilities that include newer and more diverse modalities of communication to be built into human/agent systems. To do so, we must explore embodiment in much greater depth. At present, the robots that are adopted today look anything but *human*, in terms of appearance, forms of interaction, and behavioral patterns. Generally speaking, the adoption of robots in health care cries out for the understanding of human factors, such as perception and trust, to be combined with the technical factors of accuracy and speed. Guided design of robotic assistants, by following a set of recommendations and heuristics, can help change the current (negative) perception about robots that persists among medical and surgical staff. A key element to succeed in this task is the active participation of stakeholders and potential users in the integration of robots in the OR. This, in turn, will foster a rapport between doctors and their robotic assistants. As an example, the surgical staff can elucidate key activities and expected behaviors in the surgical arena. Once prototype systems are designed, proper training programs must be developed to assure smooth integration, defining best practices for task-sharing among hybrid doctor-robot teams, and suggesting graceful ways wherein robots could recover from errors or unexpected scenarios. The author's previous work (Wachs 2012) presented a list of requirements derived from surgical staff interviews and discussions with a number of participants over four years. This list of requirements is summarized here:

1. ***Dexterity:*** Effective handling of surgical tools, equipment, and human tissue requires high dexterity. For example, the human hand has 27 degrees of freedom (DOF) whereas most robots offer wrists with 3 DOF and tool tips with 1 DOF. In case of a robotic surgical assistant, the aforementioned configuration is sufficient for picking and handing off instruments. However, when more complex tasks are required (e.g., opening a suture bag, or knot tying) robotic hands with higher dexterity are required.
2. ***Multimodality:*** Since communication between humans is by definition multimodal, it is expected that the robotic assistant will assimilate and recognize the same form of communication. This involves proper modulation of

gestures, body language, gaze, speech, and proxemics. When users adopt more than one modality of interaction, the robot must be capable of resolving ambiguities.

3. ***Timing:*** The robot must execute actions instantly when no ambiguity exists. In cases where there is a potential for error (e.g., the command is misunderstood), previous confirmation from the operator is required. The response time desired from such systems should be similar to that exhibited by an experienced surgical assistant working in tandem with the surgeon. While the robot's response should be immediate, the motions must be smooth enough to avoid tremors or potential collisions.
4. ***Contextual Inference (mind readers):*** Experienced surgical technicians may know what will be the next surgical tool required in advance, and often before the surgeon has made an explicit request. Due to this ability to anticipate the surgeons' needs, they are often referred to as "mind readers". The same form of prediction and inference based on context is expected from a robot. When the inference is wrong, graceful recovery from mistakes is necessary.
5. ***Predictable:*** Trauma cases in the OR seem chaotic and require precise team coordination and good communication grounding for effective treatment. Robot's unexpected behavior can add confusion. Thus, it is desirable that the robot actions will be "transparent" and highly predictable to the operators to avoid potential distractions, occlusions, or interference with existing procedures.
6. ***Accuracy and Precision:*** Surgeons' requests require accurate recognition from the robot, regardless of the communication forms used to convey this request. Experienced nurses can identify surgical requests precisely with almost no false alarms. This performance level is expected from the robot, even under dynamic and cluttered conditions, such as those found in ORs. Grasping small instruments correctly and safely (e.g., sponges, gauzes, sutures, and sharps) require precise movements.
7. ***Safety:*** Established standards exist in industrial robotics for operator's safety and guidelines for robot operation are available to ensure safe operation. In addition, mechanisms such as emergency stops, proximity sensors, and physical and electronic barriers are usually in place. Nevertheless, there are no equivalent standards for tasks involving human-robot interaction in the surgical setting. Drafting such guidelines will help reduce risks related to collisions with sharp instruments, or with robot parts. Furthermore, such guidelines should also establish the proper parameter setting (e.g., operation electrical currents and voltages used by the servos in the robot), and suitable strategies for collision avoidance.

A systems-based approach is required to include these requirements with existing work environment constraints, and regulatory issues concerning patient safety. There are specific tools that can support the development of such systemic approaches. One example of such tools is OPCAT (Object-Process CASE Tool) which assists in the development of conceptual models, discussed in detail in the next section.

4.4 Conceptual modeling as a way to determine modalities of communication

4.4.1 Definition and terminology

Conceptual modeling is a process that allows the description and analysis (through simulation) of a problem in a systematic fashion, with instances, factors, and processes involved. Due to the complexity and the number of communication events occurring in the OR, the adoption of tools for modeling these processes, their relationships and how they are affected by processes' outcomes is of paramount importance (Brazen 1992; Asplin et al. 2003; Bigdelou et al. 2011; McLaughlin 2012). The conceptual system described in this section allows a qualitative assessment and potential solutions of problems concerning miscommunications in the OR. It also models scenarios including those where instruments are retained in the patient during surgery and unsafe handling of surgical instruments.

The main goal of such conceptual model is to allow a faithful representation of the dynamics and interactions of fundamental elements (processes, instances, and relations) and to enable a realistic simulation of these interactions in the surgical setting through this model. The specific goals that are accomplished through this form of modeling are validated through ground truth, expert knowledge, and/or reference points for model validation and guidelines. Subjective and objective metrics to assess the success of the model must be established as part of the modeling process. For example, in the case of the OR's team communications, the metrics are the percentage of errors in the delivery of surgical instruments, the number of incidents involving mishandling of equipment, and the retained instruments within patients following surgery.

The inputs and outputs relating the different processes are obtained through empirical observations and expert knowledge. These cues (also referred as signals) should provide enough information to ensure that the modeling objectives that are defined are adequately met. Examples of these signals are the recognition

accuracy of verbal and non-verbal requests; the delivery time of the instruments; the timestamp, type and number of the retained instrument. Determining these inputs/outputs explicitly requires an implementation phase.

The conceptual model is universal in the sense that it does not specify how the different processes should be implemented in practice. In practice, this step requires the development of effective algorithms for gesture and speech recognition; robust manipulation and classification methods for surgical instruments; safe path planning; and obstacle avoidance algorithms.

4.4.2 A visual example

The conceptual model follows the OPM (Object-process methodology) principles for modular and scalable modeling, and it is implemented using the OPCAT tool (Dori, Linchevski & Manor 2010). The example presented focuses on the OR toolset handling system activity, while capturing critical communication aspects of surgery, especially those involving communication exchanges related to the handling of surgical instruments. The key component of this model is the main function of the system being modeled, which is *OR toolset handling* (Fig. 4.2), denoted as an ellipse. The second process depicted is *Operation*, which is considered environmental (dashed ellipse). The remaining elements are objects (the rectangular boxes), and links connecting objects with one another or connecting

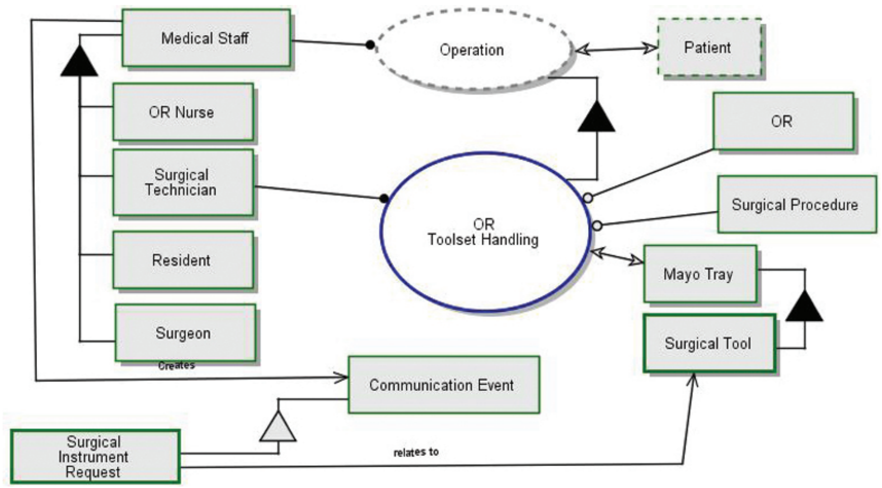


Fig. 4.2: Object-process diagram (OPD) scheme for the OR toolset handling function.

objects to processes. In this specific example, the interacting objects include the members of the surgical team since their state affects the communication events. Another element is the agent link which is a line ending with black circle at the process end. See for example the object *Medical staff* which acts as the agent for the *Operation* process. Concepts such as "Medical Staff handles Operation" is expressed by a graphic construct of the *Medical Staff* object linked with an agent link to the process *Operation*.

The schema presented in Fig. 4.1 allows visualizing key activities. For example, it shows how the patient and the surgical staff interact through an "*operation*" (process) and the *Surgical technician* interacts with the *Mayo tray* through the process "*OR Toolset Handling*". Accurate modeling of these key activities in the form of relations between objects and processes can help detect miscommunications related to instrument handling (e.g., incorrect instrument counts that can lead to retained instruments within the patient).

The remaining elements and interactions presented in the OR toolset handling process depict existing activities in the OR modeled through links and processes.

4.5 Importance of embodiment in human-machine interaction

Embodied cognition (Lakoff & Johnson 1980) is the theory that all aspects of our cognition are shaped by aspects of our body, or in other words, the nature of the human mind is mostly determined by the shape of the human body. Reasoning and decision-making are influenced by the motor system, and the physical interactions with the environment, just as bodily actions are influenced by the mind (Borghi & Cimatti 2010). This claim has been tested in a number of experiments in the following areas: visual search (Bekkering & Neggers 2001), distance perception (Balcetis & Dunning 2007), language processing (Glenberg et al. 2010), and memory (Scott, Harris & Rothe 2001).

Embodied interaction describes the way people interact cognitively and physically with information technology. This involves the way that the technology is manipulated, shared and the level of engagement that the user experiences. The emphasis of the interaction is placed mostly on the physical engagement, using hands and gaze, and the body as a whole. Examples of how this type of interaction supports cognitive processes are found in science education (Pirie & Kieren, 1994; Nemirovsky et al. 1998; Alibali, Bassok & Olseth 1999; Lakoff & Nuñez 2000), music (Leman 2007), performing arts (Mann, Janzen & Fung 2007), and gaming. Gaming consoles and new sensors such as Nintendo Wii U (Regersen 2011),

Microsoft Kinect and Leap-Motion rely fundamentally on the concept of embodied interaction to shape the gaming environment. These consoles can reliably track and recognize user movements and reflect changes in the game's environment accordingly, thus offering a more realistic experience.

Emergent motor coordination patterns in response to dynamically changing environments could result in realistic and effective hand gestures. Gestures are the result of body-environment interaction dynamics, which acts as a non-linear and time-varying system. The neural system exploits the physics of the body, and at the same time, the body dynamics shape the neural dynamics via sensory stimuli. This constitutes a fundamental property of embodiment (Brooks 1991; Pfeifer & Scheier 1999; Pfeifer & Bongard 2006). Such a model was used as the basis for bipedal motion in robotics, and can be extended to autonomous generation of a rich "human-like" variety of dynamic patterns that resemble hand gestures.

Gestures can be generated using emergent and dynamic embodied behavior resulting from simulating the effects from a combination of hardware (robot and sensors) and interconnected neural oscillators (coupled chaotic systems) (Kuniyoshi, Suzuki & Sangawa 2007). Reliable hand movements and configurations are obtained through a model of a musculoskeletal system, which resembles the human hand and arm. This model compromises a number of chaotic elements where each of these elements controls a muscle based on local sensed feedback. The chaotic elements interact through a physical body (the robotic manipulator) and the environment (sensed forces resulting from the torque, friction, and gravity). Gestures are then generated by a robotic manipulator where each actuator in the arm responds to an input signal generated by such chaotic elements.

4.6 Analyzing the performance of three cyber-physical systems designed for the operating room

In this section, the author presents three different kinds of cyber physical systems in which multimodal interaction was adopted for use with both physical and virtual (without embodiment) agents in the operating room. The main forms of interaction used are gesture and speech. The interaction took place between the surgical staff and cyber-physical agents.[2] The goal of human-robot interaction is

2 Cyber-physical agents are part of a broad class of cyber-physical systems, which can be best defined as computational elements that control some aspect of the physical environment. For example, a network of computer systems, such as PACS, would constitute a cyber-physical agent.

allow the robot to play an assistive role in performing those activities in the OR which are generally time consuming, risky, or present an increased risk to the rate of infection. The author exemplifies how collaboration with robots or cyber-physical systems has the potential to improve both patient-care outcome (both quantitatively and qualitatively) by adding such technologies to the surgical setting.

The systems presented in the following subsections are meant to: (a) support the interaction with picture archiving and communications system (PACS) in the OR; (b) enable collaboration during the surgical setting by handing surgical instruments as required by the surgeons; and (c) augment and extend surgical training through cyber-embodiment. Each of these systems is presented below.

4.6.1 Gestix

Browsing, navigation, and visual analysis of PACS images during surgery is cumbersome and relies on a variable chain of commands. When the surgeons want to access medical images in an electronic form using PACS, the assistance of a surgical nurse or technician is required. This is due to the fact that the surgeon cannot touch the PACS station without "breaching in asepsis" (a technical term that means contaminating the sterile zone) and potentially spreading serious infections. Therefore, navigation instructions (e.g., as "zoom-in," "zoom-out," "rotate," and "browse.") are delegated to the surgical support staff. While such instructions are critical for protecting the patient from infection, they can result in additional delays, miscommunications, and potential risks to the patient when for example a surgeon may be forced to stop what they are doing and take over the navigation task for a support team member who may be unavailable at that moment to perform the image retrieval task.

Obviously, one possible way to avoid these negative effects is to enable the surgeon to interact directly with the visual information through a touch-free modality. In this vein, hand gestures offer an intuitive form of interaction that is totally sterile and natural to the human operator. This interaction form allows the surgeon to remain within the operative field, while allowing them to use gestures to control the PAC system. While this approach was first proposed in 2004 (Graetzel et al. 2004), it was not introduced in the operating room until 2007 and given the name *Gestix* (Wachs et al. 2007). Even then, it was introduced in a very limited fashion: specific procedures with a limited period of interaction were allowed. An example of this application is in nonsurgical biopsies. This type of biopsy requires "frozen sections" analysis, which require about 20 minutes to complete.

During this analysis, the surgical staff discusses, re-plans (e.g., opt for taking a biopsy in a different region instead) and manipulates MRI images within the PAC system. This process does not jeopardize patient safety or incur additional delays, namely because *Gestix* allows surgeons to use hand gestures to interact with the PAC system for image navigation, manipulation, and browsing without having to touch the PACS station.

There are constraints placed on the Gestix-assisted surgeon nevertheless. Specifically, the gestures must be performed within a specific region of interaction (in other words, a specific physical location within the operating room), and the users are constrained to use only those gestures that are part of the lexicon already built into the user interface by the system designer. Computer vision tracking and recognition algorithms were developed to make sense of the gestural interaction. The recognized gestures, in turn, are converted into operational commands for image navigation and manipulation, such as “zoom-in,” “zoom-out,” “rotate,” and “browse.” Since *Gestix* relies mostly on optical information for gesture recognition, occlusions, illumination, clutter and other similar problems are likely to compromise the system’s performance.

In the last decade, speech recognition has been suggested as a potential solution to maintain the sterility in the OR and allow for the surgeon’s independent system operation. However, voice recognition interfaces have not gained much traction when used as *single* modality of interaction. The reason is that the OR tends to be a very noisy environment, due to equipment beeps and alerts, staff members conversing with one another, and other reasons. In addition, the requirement of wearing masks further compromises speech recognition accuracy rates because voice commands issued by a member of the surgical staff may sound muffled and unclear underneath those surgical masks, and are likewise affected by noise. In fact, a much research has been conducted on the acceptable noise levels in the clinical setting, and their effects on patients’ safety[3] (Kahn et al. 1998; Hickam et al. 2003; Darcy Hancock & Ware 2008; Choiniere 2011). Also, it is also not uncommon for operating rooms to be exposed to excessive noise levels due to the use of specific surgical instruments, especially those used to perform orthopedic procedures (Ginsberg et al. (2013). In view of these practical considerations, one must weigh whether or not to use speech recognition in the design of surgical robots.

3 Noise levels in several mid-Atlantic region neonatal intensive care units (NICUs) were found to be above the American Academy of Pediatrics, the recommended impulse maximum of 65 dB, and the standard established by the Environmental Protection Agency.

Gestix has gotten a boost from the rapid development of motion controllers and motion sensors that make gesture-based robot commands doable. For example, the advent of the Kinect and the Leap motion depth cameras along with the MIO wristband sensors have enabled a vast development of hand gesture-based recognition systems in recent years. The advantages of these devices are that they have been successfully tested in the surgical setting, and are affordable and easy to deploy (Kirmizibayrak et al. 2011; Gallo 2013; O'Hara et al. 2014). In the coming years it is expected that his type of technology will lead to significant development of an entire class of other gesture-based interfaces for navigation and manipulation for PACS in the OR.

A word of caution is still advisable: while this technology seems promising, there are a number of technical and conceptual limitations involved with its use. From the technical stand point, occlusions, number of human operators, proxemics and tracking reliability are still challenging issues. From the conceptual standpoint, however, problems related to human patterns of behavior are much more difficult to solve than technical ones. For example, how can the interface "infer" that the surgeon's gesture is being performed with the intention of interacting with the system (an "intentional" gesture), as opposed to a gesture that is simply meant for communicating an idea to the surgical staff (an "unintentional" gesture)? Similarly, how do we know when the gesture performed is part of the surgical task (making an incision while holding a scalpel), requesting a surgical tool (open palm for hemostat), or an actual navigation command directed to the PAC system? No doubt, such communicative ambiguities are related to the problem of contextual inference.

Some of these concerns mentioned above have been addressed by the Gestix II system developed by Jacob and Wachs (2013), where contextual inference is computed based on environmental and visual cues. The context is extracted from view-dependent anthropometric information, and task related information (e.g., the current phase in the surgery). Knowing what the surgeon is doing at a specific point in time during the surgery is a good proxy to infer what would be their future operational needs. These include visualization related commands, and manipulation and navigation operations of the medical images. Being able to infer intention and action from context leads to a significant reduction in the number of false positives in command recognition. This means that the system can precisely discriminate those gestures that are not meant to be used for operational control, whereas before, those movements were mistakenly recognized as intended commands. While the use of speech as a *single* modality may not be suitable for the surgical setting (due to the excessive noise and other factors that are mentioned above), a combination of gesture and speech may support the surgical

Fig. 4.3: *Gestix* operated by a surgeon in the operating room at the washington hospital center.

task more effectively than using each modality by itself. The reason for this is that multimodal interaction provides a healthy form of redundancy, which is a key factor when recognition (of a surgeon's command) based on a single modality may be ambiguous. Several aspects of multimodal interaction are explored in the next section, which describes another kind of cyber-physical system.

4.6.2 Gestonurse

Delivery and retrieval of surgical instruments constitutes one of the main tasks assigned to the surgical scrub nurse in the operating room. This is a repetitive and monotonous task, which takes most of the attention of the surgical nurse. The task it is not necessarily a difficult one, however "high situation awareness" (a term that is often used in aviation and other fields to mean keen perception of one's environment) is required. Thus, passing the wrong instrument can lead to unnecessary delays and mistakes, and increase the risk of surgical complications. The surgical nurse is also responsible for operating sterilizers, lights, suction machines, electrosurgical units, and diagnostic equipment, as well for holding retractors, applying sponges, or suctioning the operative site. However, their main responsibility is delivery, retrieval, and tracking the use of surgical instruments.

Initial attempts to automate this activity of passing along surgical instruments, as part of a larger effort involving the development of a robotic scrub nurse, relied on single modalities. For example, spoken commands were used to request the surgical instruments from the robotic nurse (Kochan 2005; Treat et al. 2006; Gilbert, Turner & Marchessault 2007). The spoken commands where, in turn, converted into commands representing the set of surgical instruments. Such commands are compromised, however, by environmental noise which affects the performance of a speech recognition system (Ginsberg et al. 2013).

A recent systematic study conducted by the author and his colleagues at Indiana University School of Medicine involving empirical observations of how surgical teams communicate with one another in the OR with regard to the use and management of surgical instruments led to initial findings about this task. These study findings indicate that the communication between the main surgeon and the surgical technician/or surgical nurse is comprised mainly of gestures, speech and proxemics (Jacob et al. 2012; 2013b). These findings dictated the minimum requirements in which a robotic scrub nurse should communicate. *Gestonurse* (Jacob et al. 2012a; Jacob, Li & Wachs 2012; 2013b) is the first multimodal robotic scrub nurse developed at Purdue with such multimodal capabilities. This system can pick surgical instruments, and retrieve and count surgical instruments within the operative site. The author and his research group have been studying *Gestonurse* to see how effective this robotic assistant is at performing surgical instruments delivery (see Fig. 4.1). A robot with a multimodal interface and robust recognition algorithms can reliably resemble the surgeon-nurse work in tandem. Such a robot could potentially take over some of the tedious tasks commonly performed by surgical technicians.

This is how the robotic surgical task flows: the main surgeon requests the surgical instruments based on their needs during the surgical procedure; those instruments are then immediately handed off to the surgeon by a robotic manipulator. The surgeon uses one or more instruments at a time. The instruments that are no longer required during the surgical procedure are left to one side of the operative site. In turn, the robot retrieves the instruments that are no longer required.

Surgical instrument requests are transmitted through two main communication forms: explicit and implicit. The explicit form is verbal or physical (e.g., gestures), and the implicit form is based on inference. This type of inference is most common in surgery. The difference however between human-human surgical interactions, and those that are assisted by surgical robots is that unlike the surgical technician who can predict the type of instrument and when to deliver it (which is why they are called “mind readers,” as mentioned above) the surgical robot cannot easily pick up on inferences. As such, *Gestonurse*, for example, relies on the surgeon’s *explicit*

communication. It is able to recognize spoken commands using speech recognition algorithms, as well as gestures (both static and dynamic) which serve as the vehicle to request the instruments. The set of gestures used for the requests are referred as the "gesture lexicon." This lexicon includes poses and movements, which are naturally performed by surgeons in standard surgeries while other gestures require a bit of training. For example, open-palm indicates the need of a hemostat, which is very intuitive; or two fingers opened in "V" shape representing "scissors". In contrast, those gestures that are not naturally used by surgeons require a training period for both the robot and the surgeon so that robot can recognize those gestures and what they mean. The duration of this training depends on the size of the lexicons and the surgeons' familiarity with the gestures they must use in communicating with the robot.

This problem of communicating with the surgical robot does not exist in systems which are solely speech-driven, since the instruments' names are fairly standard. Multimodal communications, however, pose challenges since gestures are not entirely uniform, and thus their association with a surgical instrument in the act of making a request for that particular instrument is not necessarily standard within a particular culture. Yet, in spite of the obstacles posed by gesture communication, Jacob and Wachs (2013) reported that the required amount of time for robots to learn how to recognize and correctly interpret gestures is not excessive, and the increase in performance certainly outweighs the time it takes to train the robot.

There are, however, two hurdles that serve as impediments to the adoption of robotic multimodal robots in the OR. The first is related to health and safety risks entailed in the use of automation in proximity to a surgeon and patient. For example, the when a robot passes a sharp instrument at the time a nurse moves their hand to request the instrument. This can cause to injuries and can lead to infections of the nurse and patient. Therefore reactive obstacle avoidance, dynamic planning, and on-line learning are some of the key requirements to assure a safe environment for the robot-human surgical team. The second problem is related to having the robot predict the instrument required by the surgeon. Algorithms can be used to "learn" patterns of behavior based on hundreds of surgeries observed, and act according to new patterns that resemble in some way those learnt previously. While this approach can be successful for established and routinely performed surgical procedures, it can hardly be applicable to surgeries that were not planned in advance (such as trauma surgeries) or, alternatively, those procedures where unexpected surgical complications occur. In both cases the sequence of instruments cannot be established beforehand. Developing mechanisms for prediction that can dynamically adjust to the existing scenario is required to avoid chaos in the OR when such unpredicted events occur.

As a final note on this system, clinicians, surgeons, and surgical technicians have shown interest in having this type of cybernetic solution as part of the surgical setting, subject of course to suitable solutions to the kinds of problems mentioned above. In addition, accurate and fast delivery (compared to that of a surgical assistant) have been mentioned likewise as a desirable feature of surgical robots. Furthermore, a compact, lightweight and fast configurable system will allow mobility between the ORs, rather than allocating specific rooms for the robots. Future desired capabilities include enabling the robot to conduct more complex supporting tasks, or even perform parts of the surgery that are of a more routine nature. Such capabilities will be one of the features discussed in the section below.

4.6.3 Telementoring

Treating trauma injuries effectively and promptly requires the kinds of surgical skills and proficiency found mainly in the major teaching hospitals in the US. Unfortunately, such skills are not usually found in the smaller hospitals found in rural America. This so because small hospitals often lack the surgical expertise

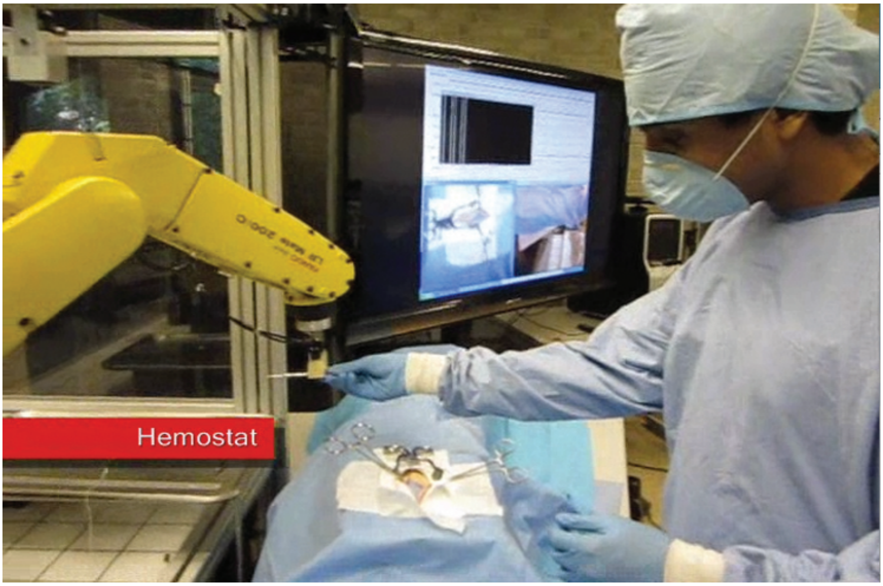


Fig. 4.4: *Gestonurse* delivers surgical instruments to the surgeon as required.

required to handle traumatic injuries (Shively & Shively 2005). Borgstrom (2011) have pointed out that in the last few years it has been widely reported that rural hospitals are lacking the number and type of surgical expertise necessary to treat the conditions presented by the populations in rural regions. This population is overall sicker, older, poorer, and less well educated than their counterparts in the cosmopolitan regions. Furthermore, percentages of infant mortality and injury-related mortality are greater in rural areas. Most rural general surgeons do not have the necessary training to perform trauma procedures, and the demand for surgeons is expected to rise by more than 30% in the next 15 years, exacerbating the risks to patient safety even more. Depending on the surgery type, 15 to 100 surgeries are necessary to reach the plateau of the learning curve (Zhou et al. 2012). This is the number of procedures required for a trainee to master a subspecialty and achieve a low complication rate (Wang 2011). A similar situation is found in the battlefield where field hospitals need to treat blast and fragmentation injuries requiring appropriate care from a surgical expert, such as a neurosurgeon, who may not be physically available in the field.

In both the case of the patient confined in a rural hospital or the patient stuck on the battlefield, commuting to a level 1 trauma center may not be advisable since it could jeopardize the patient's life, in addition to incurring additional costs and logistical difficulties. Nevertheless, delays in treatment are found to be a contributing factor in trauma-related deaths (Abolhoda 1997; Manlulu 2004). In such cases, the patient needs to be treated at the point of care with limited surgical resources, though lacking the necessary expertise for effective treatment. Real-time instruction from a specialist surgeon is required for appropriate and immediate medical care in this austere environment. This specialist could walk the frontline surgeon through the surgical procedure, which the mentee surgeon may not have seen in the past such as a craniectomy. In this context, *telementoring* can be a key component in the optimal treatment at the point of care, whether this occurs at a rural hospital or a forward operating base[4] in the battlefield.

Telementoring involves procedural guidance of a trainee (mentee) surgeon by an expert surgeon (mentor) from afar using information technology and telecommunication. This method has been shown to be practical for providing real-time instruction, guidance, and consultation remotely through, audio, video and haptics. Chebbi, Lazaroff and Liu (2007) show how haptics, as a form of non-verbal communication involving touch, is used to assist surgeons in performing an unfamiliar procedure by using "force feedback". The way this is done is by

4 A forward operating base (FOB) is a military base used to support tactical operations in a secured forward military position.

having the video feedback presented to the mentee on a nearby HD display or through a high quality telestrator. A telestrator is a device that allows the remote mentor to draw, annotate, sketch and point over a video image displayed to the mentee remotely. While haptics have been used in minimally invasive surgery (MIS) in concert with audio and video instruction, this is not the case in trauma surgery where there is no effective way to convey tactile information to the expert surgeon. In MIS, force feedback can help guiding the laparoscope by the mentor, and serve as an additional form of instruction during an MIS procedure. This is not applicable, however, to open surgery for the simple reason that any external force exerted on the trainee's hand can affect the precision of the surgical movement leading to catastrophic results.

In addition to audio, video and haptics another key component in surgical instruction is the use of gestures. These gestures are also referred as *surgical instructional gestures (SIGs)* (Wachs & Gomez 2013), and occur throughout the mentor-trainee surgical training. Conveying these gestures through telementoring is a particularly challenging task and a virgin area of research. The ability to generate meaningful gestures through agents/robots is referred to as *embodiment* in the human robot interaction (HRI) scientific community. Through embodiment, the mentor would convey gestural instruction to the mentee at the remote site. In such a scenario, the gestures would be produced by a robot, which would be

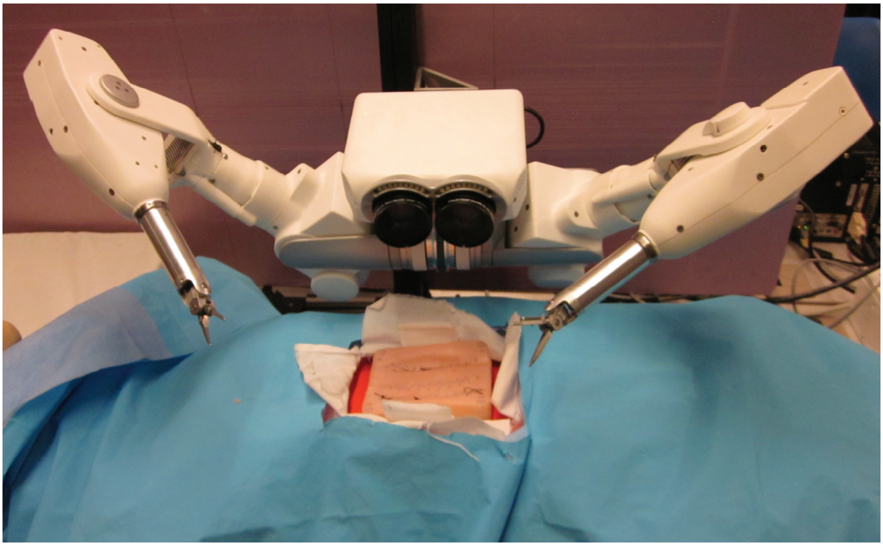


Fig. 4.5: The taurus robot performing surgical instructional gestures (SIGs).

controlled by the expert surgeon. Telementoring in combination with embodiment through surgical robots (see Fig. 4.2) may present pedagogical benefits in terms of better and faster remote surgical training, and comparable to performances exhibited by mentors and mentees when they are co-located in the same physical space.

Recent research focused on the effective of use visual communication to improve the sense of co-presence in telementoring systems. For example, augmented reality was implemented on tablets or see-through-displays to display mentors' annotations over the patients' anatomy. This form of cyber-interaction allows delivering spoken and visual cues about the surgical action blended with annotations over the operative site. Other innovative approaches involve projecting these annotations directly on the patient (e.g., through laser technology) (Ereso et al. 2010), or displaying a projection of the hand movements of the expert surgeon on the remote display (Shenai et al. 2011).

4.7 Discussion and conclusions

In the past decade, information technology (IT) has had a major impact on health care, resulting in marked improvements in patient care from diagnoses to successful treatment. IT has likewise led to overall organizational improvements in the healthcare system from access to patients files to extracting information from huge pharmacological and histological databases that are relevant to patient care. The inclusion of cybernetics however, as opposed to other IT technologies, has continually faced additional challenges due to regulatory, safety and societal concerns that have not yet been fully addressed. This is surprising considering that the cyber-based solutions have been shown to provide direct improvements in health care process and outcomes, especially those solutions that enhance the practitioner's precision and timing. As an example, the reader can refer to objective and economic benefits directly linked to the introduction of surgical robotics into the operation theater. Nevertheless, the healthcare community seems hesitant to integrate these technologies for a number of reasons.

This chapter discusses some of the societal and technical challenges involved in the adoption of robots in health care and their potential for improving patient outcomes. For example, miscommunications was indicated as one of the leading causes for mistakes in the operating room, leading to increasing risks of mortality and morbidity. In this context, a cybernetic solution can take the form of a robotic assistant that can interpret multimodal communications among the surgical team and act according to their expectations. For example, a robotic assistant

could recognize spoken and nonverbal commands, detect and deliver surgical instruments, and assist the leading surgeon through the procedure as required. To achieve this goal, significant improvements are necessary for sense-making, prediction, and interaction in such intricate environments. One of the challenges stressed in this chapter has to do with the social acceptance and trust of these robotic assistants, and how well they can be integrated into existing surgical teams. Positive perception and greater trust is attained as a response to increasing success with the use of robotic agents in the medical setting. This can only occur once the technical roadblocks are cleared, such as the lack of accuracy, speed, and flexibility to adjust to uncontrolled conditions (e.g., unfixed lighting, clutter, or deviation from a standard procedure) which are commonly found in healthcare environments.

In order to engage these cybernetic solutions in the most meaningful ways, it is necessary to understand and quantify accurately the type of processes and the nature of interactions among these processes in the relevant healthcare domain (e.g., operating room). There are a number of approaches to model the complex interactions existing among the agents in a dynamic setting. Through this chapter, we proposed the OPM as an attractive modeling alternative, which offers flexibility and easiness of representation. This model allows straightforward process visualization, and analysis of their affects on the interacting entities. The modeling process involves the participation of domain experts and stake-holders (e.g., surgeons, nurses, surgical technicians and human-factors engineers) from its conception all the way to the final design and testing. Once the model is completed and validated through numerous direct observations, sketches, records and video footage, each process is examined in search for existing pitfalls, mistakes and potential improvements. The final step on this validation is to cross-compare the existing capabilities to those offered by the cybernetic agent. Then, substitution implications are analyzed towards the mentioned capabilities to assure that no negative effects would be introduced in the healthcare setting as a result of changes that may occur during this process.

An additional point discussed in this chapter involves ways for substituting physical expression (intrinsic in human inter-personal communication) by artificial artifacts generated through the robot. This feature is dubbed “embodiment,” and involves all the forms of expressions conveyed through the human body. Embodiment theory is a particular “hot” area of research within the human-robot interaction field, which includes computer scientists, engineers and psychologists. Venues where these topics are discussed and studied are conferences such as the ACM/IEEE International Conference on Human-Robot Interaction, and journals such as the Journal of Human-Robot Interaction.

The chapter concludes by discussing three different applications where robotics and intelligent agents were evaluated in the healthcare setting and have shown their potential impact. The first application is *Gestix*, which allows the surgeon to browse medical images just by hand movements and static hand postures. Since the introduction of this system, several others have followed this path and have offered the potential user a touchless form of interaction with medical records and PACS systems (Kirmizibayrak et al. 2011; Gallo 2013; O'Hara et al. 2014). In spite of this overwhelming surge of applications, key issues must be addressed, such as how to track reliable hand gestures with multiple users under dynamic illumination and through occlusions. Other critical questions include how to disambiguate control actions from the surgical movements necessary during surgery. We presented some results tackling this problem; nevertheless more work needs to be devoted to address questions such as scalability and design of the gesture lexicon for effective interaction between surgeons and robots.

The second application demonstrates the implementation of a robotic assistant for the operating room that can understand multimodal interaction. The assistant's main role is to deliver surgical instruments as required by the lead surgeon. The key concept introduced through this application is the idea of surgical co-robots, meaning that the robot works together with the surgeon, rather than being teleoperated by them (as is conventionally done). Through the implementation of this concept, challenges related to the prediction of the next phase of surgery, proxemics recognition and safety standards have been discussed. Those challenges must be addressed properly before any type of robotic assistant will be allowed to participate and support the surgical team during surgery.

The last case study involves a telementoring system. This system is meant to be used to instruct/guide a mentee surgeon (non-expert surgeon or a trainee surgeon) to conduct surgery remotely, supported by cybernetics and information technology. In this context, an important contribution discussed will be incorporating gesture production through embodiment embedded within a robotic assistant. Preliminary work has been conducted to determine the fundamental set of gestures involved in surgical training (also referred as SIG's). The ability to reproduce these instructional gestures will be a feature desired in future telementoring systems.

In addition, regardless of the robotic system used, speech must be explored as an integral feature of human-robot interaction so that it can be optimally used in the OR notwithstanding the noisy background and other factors that may compromise speech recognition accuracy rates.

All in all, the introduction of surgical robots in the surgical arena (as assistants rather than autonomous agents) will have sociological and technological

implications that will aid in the transformation of health care to better serve humankind. To assure that those changes will lead to increased patient safety and overall better outcomes for all, key challenges must first be addressed. Once those challenges are addressed the next generation of multimodal robots will play a constructive role in bringing about enhanced patient care.

Acknowledgments

This publication was made possible by a NPRP grant #6-449-2-181 from the Qatar National Research Fund (a member of The Qatar Foundation), the US Air Force Office of Scientific Research (AFOSR) Young Investigator Research Program (YIP 2013), under grant number FA9550-13-1-0141, and this work was supported by the Office of the Assistant Secretary of Defense for Health Affairs under Award No. W81XWH-14-1-0042. Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the Department of Defense. The statements made herein are solely the responsibility of the author.

References

Abolhoda, A., Livingston, D. H., Donahoo, J. S. & Allen, K. (1997) 'Diagnostic and therapeutic video assisted thoracic surgery (VATS) following chest trauma', *European Journal Cardio-Thoracic Surgery*, 12(3):356–360.

Alibali, M. W., Bassok, M., Solomon, K. O., Syc, S. E. & Goldin-Meadow, S. (1999) 'Illuminating mental representations through speech and gesture', *Psychological Science*, 10(4):327–333.

Asplin, B. R., Magid, D. J., Rhodes, K. V., Solberg, L. I., Lurie, N. & Camargo Jr, C. A. (2003) 'A conceptual model of emergency department crowding', *Annals of emergency medicine*, 42(2):173–180.

Balcetis, E. & Dunning, D. (2007) 'Cognitive dissonance and the perception of natural environments', *Psychological Science,* 18(10):917–921.

Bekkering, H. & Neggers, S. F. W. (2001) 'Visual search is modulated by action intentions', *Psychological Science,* 13(4):370–374.

Bigdelou, A., Sterner, T., Wiesner, S., Wendler, T., Matthes, F. & Navab, N. (2011) OR specific domain model for usability evaluations of intra-operative systems. In *Information Processing in Computer-Assisted Interventions,* (pp. 25–35). Springer Berlin Heidelberg.

Brazen, L. (1992) 'The difference between conceptual models, practice models', *AORN Journal*, 56(5):840–844.

Brooks, R. A, (1991) 'Intelligence without representation', *Artificial Intelligence*, 47(1–3):139–159.

Borghi, A. M. & Cimatti, F. (2010) 'Embodied cognition and beyond: acting and sensing the body', *Neuropsychologia,* 48(3):763–773.

Borgstrom, D. C. (2011) Surgery for rural America. In *Rural Surgery,* (pp. 5–6). Springer: Berlin Heidelberg.

Cao, C. G. & Taylor, H. (2004) 'Effects of new technology on the operating room team. Tufts University of Medford, MA, Department of Mechanical Engineering.
Carthey, J., de Laval, M. R., Wright, D. J., Farewell, V. T. & Reason, J. T. (2003) 'Behavioural markers of surgical excellence', *Safety Science*, 41:409–425.
Chebbi, B., Lazaroff, D. & Liu, P. (2007) 'Collaborative virtual haptic environment for surgical training and tele-mentoring', *International Journal of Robotics and Automation*, 22(1):69–78, 2007.
Choiniere, D. B. (2010) 'The effects of hospital noise', *Nursing administration quarterly*, 34(4):327–333.
Cunningham, S., Chellali, A., Jaffre, I., Classe, J. & Cao, C. G. (2013) 'Effects of experience and workplace culture in human-robot team interaction in robotic surgery: a case study', *International Journal of Social Robotics*, 5(1):75–88.
Darcy, A. E., Hancock, L. E. & Ware, E. J. (2008) 'A descriptive study of noise in the neonatal intensive care unit ambient levels and perceptions of contributing factors', *Advances in Neonatal Care*, 8(3):165–175.
Dori, D., Linchevski, C. & Manor, R. (2010) OPCAT – A software environment for object-process methodology based conceptual modelling of complex systems. *Proceedings of the 1st International Conference on Modelling and Management of Engineering Processes*, University of Cambridge, Cambridge, UK, Heisig, P., Clarkson, J. & Vajna, S. (Eds.), pp. 147–151.
Egorova, N. N., Moskowitz, A., Gelijns, A., Egorova, N. N., Moskowitz, A., Gelijns, A., Weinberg, A., Curty, J., Rabin-Fastman, B., Kaplan, H., Cooper, M., Fowler, D., Emond, J. C. & Greco, G. (2008) 'Managing the prevention of retained surgical instruments. What is the value of counting?', *Annals of Surgery*, 247(1):13–8.
Ereso, A. Q., Garcia, P., Tseng, E., Gauger, G., Kim, H., Dua, M. M. & Guy, T. S. (2010) 'Live transference of surgical subspecialty skills using telerobotic proctoring to remote general surgeons', *Journal of the American College of Surgeons*, 211(3):400–411.
Firth-Cozens, J. (2004) 'Why communication fails in the operating room', *Quality and Safety Health Care*, 13(5):327.
Gallo, L. (2013) 'A study on the degrees of freedom in touchless interaction'. In *SIGGRAPH Asia 2013 Technical Briefs,* (p. 28). ACM.
Gilbert, G., Turner, T. & Marchessault, R. (2007) Army medical robotics research. Army Medical Research and Materiel Command, Fort Derrick, MD, Telemedicine and Advance Research Center.
Ginsberg, S. H., Pantin, E., Kraidin, J., Solina, A., Panjwani, S. & Yang, G. (2013) 'Noise levels in modern operating rooms during surgery', *Journal of cardiothoracic and vascular anesthesia*, 27(3)528–530.
Glenberg, A. M., Havas, D., Becher, R. & Rinck, M. (2005). Grounding language in bodily states: The case for emotion. In Pecher D. & Zwaan R. A. (Eds.), *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking,* (pp. 115–128). Cambridge: Cambridge University Press.
Graetzel, C., Fong, T. W., Grange, S. & Baur, C. (2004) 'A non-contact mouse for surgeon-computer interaction', *Technology and Health Care*, 12(3):245–257.
Halverson, A. L., Casey, J. T., Andersson, J., Anderson, K., Park, C., Rademaker, A. W. & Moorman, D. (2011) 'Communication failure in the operating room', *Surgery*, 149(3), 305–310.
Hickam, D. H., Severance, S., Feldstein, A., Ray, L., Gorman, P., Schuldheis, S., et al. (2003) The effect of health care working conditions on patient safety: summary. 2003 Mar. In *AHRQ*

Evidence Report Summaries. Rockville (MD): Agency for Healthcare Research and Quality (US); 1998–2005.

Jacob, M., Li, Y. T., Akingba, G. & Wachs, J. P. (2012) 'Gestonurse: a robotic surgical nurse for handling surgical instruments in the operating room', *Journal of Robotic Surgery*, 6(1):53–63.

Jacob, M. G., Li, Y. T. & Wachs, J. P. (2012) Gestonurse: a multimodal robotic scrub nurse. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction,* (pp. 153–154). ACM.

Jacob, M. G. & Wachs, J. P. (2014). 'Context-based hand gesture recognition for the operating room', *Pattern Recognition Letters*, 36, 196–203.

Jacob, M. G., Li, Y. T., Akingba, G. A. & Wachs, J. P. (2013) 'Collaboration with a robotic scrub nurse', *Communications of the ACM*, 56(5):68–75.

Kahn, D. M., Cook, T. E., Carlisle, C. C., Nelson, D. L., Kramer, N. R. & Millman, R. P. (1998) 'Identification and modification of environmental noise in an ICU setting', *CHEST Journal*, 114(2):535–540.

Kirmizibayrak, C., Radeva, N., Wakid, M., Philbeck, J., Sibert, J. & Hahn, J. (2011) Evaluation of gesture based interfaces for medical volume visualization tasks. In *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry,* (pp. 69–74). ACM.

Kochan, A. (2005) 'Scalpel please, robot: Penelope's debut in the operating room', *Industrial Robot: An International Journal*, 32(6):449–451.

Kohn, L., Corrigan, J., Donaldson, M. (1999) *To err is human: building a safer health system. Institute of Medicine:* Washington, DC.

Kuniyoshi, Y., Suzuki, S. & Sangawa, S. (2007) Emergence, exploration and learning of embodied behavior. In *Robotics Research,* (pp. 473–486). Springer: Berlin, Heidelberg.

Lakoff, G. & Johnson, M. (1980) *Metaphors We Live By*. University of Chicago Press. ISBN 0-226-46801-1.

Lakoff, G. & Nuñez, R. E. (2000) '*Where Mathematics Comes from: How the Embodied Mind Brings Mathematics Into Being*'. Basic Books: New York.

Leman, M. (2008) *Embodied Music: Cognition and Mediation Technology*. MIT Press.

Li, Y. T., Jacob, M., Akingba, G. & Wachs, J. P. (2013) 'A cyber-physical management system for delivering and monitoring surgical instruments in the OR', *Surgical Innovation*, 20(4), 377–384.

Lingard, L., Espin, S., Whyte, S., Regehr, G., Baker, G. R., Reznick, R., Bohnen, J., Orser, B., Doran, D., Grober, E. (2004) 'Communication failures in the operating room: an observational classification of recurrent types and effects', *Quality and Safety in Health Care*, 13:330–334.

McLaughlin, M. M. (2012) A model to evaluate efficiency in operating room processes. Doctoral dissertation. The University of Michigan.

Manlulu, A. V., Lee, T. W., Thung, K. H. Wong, R. & Yim, A. P. (2004) 'Current indications and results of VATS in the evaluation and management of hemodynamically stable thoracic injuries', *European Journal Cardio-Thoracic Surgery*, 25(6):1048–1053.

Mann, S., Janzen, R., Fung, J. (2007) Non-electrophonic cyborg instruments: playing on everyday things as if the whole world were one giant musical instrument. In *Proceedings of the 15th Annual ACM International Conference on Multimedia*, September 23–28, Augsburg, Germany.

McCulloch, P., Mishra, A., Handa, A., Dale, T., Hirst, G. & Catchpole, K. (2009) 'The effects of aviation-style non-technical skills training on technical performance and outcome in the operating theatre', *British Medical Journal*, 18(2):109.

Mitchell, L. & Flin, R. (2008) 'Non-technical skills of the operating theatre scrub nurse: literature review,' *Journal of advanced nursing*, 63(1):15–24.

Nemirovsky, R., Tierney, C. & Wright, T. (1998) 'Body motion and graphing', *Cognition and Instruction,* 16(2):119–172.

O'Hara, K., Gonzalez, G., Sellen, A., Penney, G., Varnavas, A., Mentis, H. & Carrell, T. (2014) 'Touchless interaction in surgery', *Communications of the ACM*, 57(1):70–77.

Pfeifer, R. & Scheier C. (1999) 'Understanding Intelligence'. MIT Press.

Pfeifer, R. & Bongard, J. C. (2006) '*How the Body Shapes the Way We Think: A New View of Intelligence*. MIT Press.

Pirie, S. E. B. & Kieren, T. (1994) 'Growth in mathematical understanding: How can we characterize it and how can we represent it?', *Educational Studies in Mathematics* 26:165–190.

Regersen, A. (2011) 'Genre, technology and embodied interaction: the evolution of digital game genres and motion gaming' MedieKultur. *Journal of Media and Communication Research*, 27.

Scott, C. L., Harris, R. J., & Rothe, A. R. (2001) 'Embodied cognition through improvisation improves memory for a dramatic monologue', *Discourse Processes,* 31(3):293–305.

Shenai, M. B., Dillavou, M., Shum, C., Ross, D., Tubbs, R. S., Shih, A. & Guthrie, B. L. (2011) 'Virtual interactive presence and augmented reality (VIPAR) for remote surgical assistance', *Neurosurgery*, 68, ons200-ons207.

Shively, E. & Shively, S. (2005) 'Threats to rural surgery', *American Journal of Surgery,* 190(12):200–2005

Taylor, R. H. & Stoianovici, D. (2003) 'Medical robotics in computer-integrated surgery', *IEEE Transactions on Robotics and Automation*, 19(5).

Treat, M. R., Amory, S. E., Downey, P. E. & Taliaferro D. A. (2006) 'Initial clinical experience with a partly autonomous robotic surgical instrument server', *Surgical Endoscopy And Other Interventional Techniques,* 20(8):1310–1314.

Wachs, J. P., Stern, H., Edan, Y., Gillam, M., Feied, C., Smith, M. & Handler, J. (2007) 'Gestix: a doctor-computer sterile gesture interface for dynamic environments'. In *Soft Computing in Industrial Applications,* (pp. 30–39). Springer: Berlin, Heidelberg.

Wachs, J. P. & Gomez G. (2013) 'Telementoring systems in the operating room: a new approach in medical training', *MEDICINA (Buenos Aires)*, 73(6):539–542.

Wang, B., Lü, G., Patel, A. A., Ren, P. & Cheng, I. (2011) 'An evaluation of the learning curve for a complex surgical technique: the full endoscopic interlaminar approach for lumbar disc herniations', *Spine Journal*, 11(2):122–130

Wachs, J. P. (2012) Robot, pass me the scissors! How robots can assist us in the operating room. In Lecture Notes in Computer Science, 2012, Volume 7441, *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications,* pp. 46–57.

Webster, J. L. & Cao, C. G. (2006) 'Lowering communication barriers in operating room technology', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(4):747–758.

Zhou, M., Tse, S., Derevianko, A., Jones, D. B., Schwaitzberg, S. D. & Cao, C. G. L. (2012) 'Effect of haptic feedback in laparoscopic surgery skill acquisition', *Surgical Endoscopy*, 26(4):1128–1134.

2014 IEEE International Conference on Systems, Man, and Cybernetics
October 5-8, 2014, San Diego, CA, USA

An Augmented Reality Approach to Surgical Telementoring

Timo Loescher
School of Industrial Engineering
Purdue University
West Lafayette, IN
tloesche@purdue.edu

Shih Yu Lee
School of Electrical
and Computer Engineering
Purdue University
West Lafayette, IN
lee1549@purdue.edu

Juan P. Wachs*
School of Industrial Engineering
Purdue University
West Lafayette, IN
jpwachs@purdue.edu

***Abstract*—Optimal surgery and trauma treatment integrates different surgical skills frequently unavailable in rural/field hospitals. Telementoring can provide the missing expertise, but current systems require the trainee to focus on a nearby telestrator, fail to illustrate coming surgical steps, and give the mentor an incomplete picture of the ongoing surgery. A new telementoring system is presented that utilizes augmented reality to enhance the sense of co-presence. The system allows a mentor to add annotations to be displayed for a mentee during surgery. The annotations are displayed on a tablet held between the mentee and the surgical site as a heads-up display. As it moves, the system uses computer vision algorithms to track and align the annotations with the surgical region. Tracking is achieved through feature matching. To assess its performance, comparisons are made between SURF and SIFT detector, brute force and FLANN matchers, and hessian blob thresholds. The results show that the combination of a FLANN matcher and a SURF detector with a 1500 hessian threshold can optimize this system across scenarios of tablet movement and occlusion.**

I. Introduction

Telementoring systems benefit surgeons and medics by providing assistance from experienced mentors who are geographically separated [1]–[5]. In such systems, a remotely located mentor instructs a trainee or mentee surgeon through a surgical procedure through visual and verbal cues. The most rudimentary way to implement such a system is by using phones as a connection bridge to have the mentor verbally instruct the mentee [6]. The main limitation of using only verbal communication is that such a system limits the ability of the mentor and the mentee to share visual information. This information sharing is key to the completion of the procedure. Indicating the correct position of incisions and the placement of other surgical instruments allows for a more natural form of communication. Both visual and spoken interaction is necessary in the context of surgery. However, the flow of the surgery should not be interrupted by the surgeon's interaction with the system or focus shifts caused by the system. For this reason, obtrusive interfaces based on telestration are not suitable [7]. This paper discusses a system that offers:

1) An augmented reality interface for the mentee which displays the mentor's annotations in near real-time.
2) An algorithm to track and update the annotations on the patient's anatomy throughout the surgery.

The rest of the paper is organized as follows: first, the background of telementoring is presented along with the main gaps that this technology currently faces. Next, the architecture of the mentor/mentee system is discussed. Then, an evaluation of the core set of feature trackers and tracking accuracy performance is presented. Finally, the implications of the results are presented and the paper concludes with a summary and a discussion of directions for future work.

II. Related work

Telementoring is described as the assistance of one or more mentors on a task through verbal, tactile, and visual cues from a remote location. This remote instruction is commonly used in training and educational environments [8]–[10]. One area where recent focus has shifted regarding telementoring is in healthcare, specifically in surgical operating environments. It is not uncommon that surgeons are in scenarios when they could benefit from a subspecialist's expertise. Research has shown the benefits of the visual access to and of the remote proctoring of surgeries [1]–[3], as well as the potential for telementoring to improve minimally invasive surgery through remote video-assisted instruction [4], [5].

A newer branch for telementoring in surgery regards the utilization of visual assistance. Dixon et al. [11] looked at the effects of augmented reality on telementoring success with regard to visual attention. This discovery showed that introducing annotations such as anatomical contours to endoscopic surgeons improved accuracy, albeit at a cost to cognitive attentional resources. As this paper continues, we use the term augmented reality as defined by Augestad et al. to describe "the addition of annotations to a viewport to augment the viewer's visual information" [12].

Augmented reality allows for the real time observation of desired data or critical information in a three-dimensional environment without task interruption. In the context of surgery, this includes the monitoring of vitals and deliberation with radiological scans during an operation, both of which distract the surgeon from the primary task [13], [14]. Data has shown that such distraction in surgical tasks can be common and

The two first authors contributed equally to this work.

This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs under Award No. W81XWH-14-1-0042. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the Department of Defense.

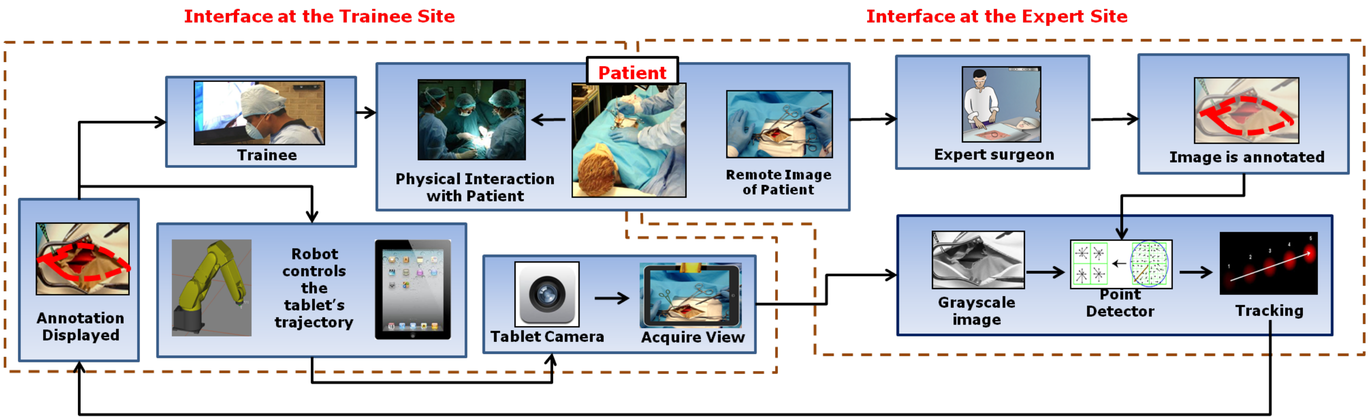


Fig. 1: System architecture

may lead to detrimental effects [15]. Even though augmented reality can bring positive assistance to surgical sites, recent research showed a disconnect from the surgical workflow due to obstacles presented by the implementations. The systems are often only beneficial for planning purposes, as they are bulky [14], their displays do not adjust over time with the region of interest [16], or the system runs too slowly to be of any use [17]. Recently, a number of applications have been developed for tablet usage during surgery [18]–[23]. They have been implemented in the context of education, navigation, and image reconstruction. However, none of these applications were used for telementoring. This paper presents a surgical telementoring system that addresses these issues. The following section discusses the design of a tablet-computer system and the evaluation of its algorithmic parameters.

III. Methods

A. System Design

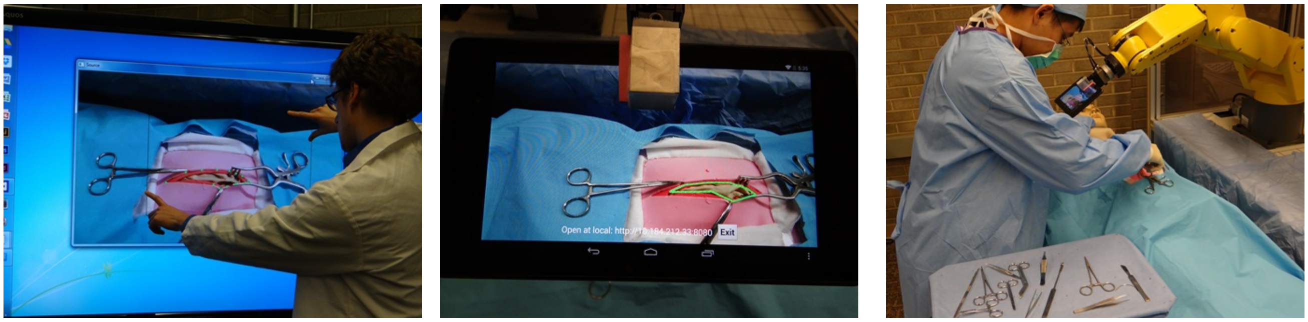
The architecture of the designed system is shown in Fig. 1. Physically, the surgical mentee stands at the patient's side when performing surgery. The mentee looks *through* a tablet screen at a real-time video feed from the rear facing camera directed at the surgical site. This affords the sense of looking through a window at the patient. The tablet is held by a robotic arm to allow the physician to move around without compromising his field of view. If the surgeon just wants a to take a glance a technologist or assistant can manually hold the tablet in his/her field of view; however, the human hand is more prone to movements and therefore may have adverse effects on the system's tracking.

At the other remote site, a mentor, who is another surgeon, is accessing the surgical view from the Internet delivered by the camera on the tablet remotely. The mentor's computer displays the video feed from the tablet for monitoring purposes. When the mentee needs guidance, the off-site mentor selects a surgical region from the tablet feed and annotates the image while the system freezes the frame, as shown in Fig. 2a. This region of interest selected by the mentor is what is referred to as the *template* throughout the paper. The annotations added by the mentor might conceptually take the form of text strings, sketches, radiology imaging overlays, or locational highlighting for tool placement. In this system, adding the annotations consists of creating a polygon on the surgical region, as well as adding strings of text (e.g. "incision", "closure"). As soon as the template is selected by the mentor, the mentor's host computer immediately begins detecting and tracking that template in incoming images. As the annotations are completed, the mentee surgeon can see the mentor's notes on the annotated window the tablet provides (Fig. 2b). Then, he/she can use those annotations by looking through the tablet while working, as displayed in Fig. 2c. This continues until the mentee no longer needs the annotations, at which point the annotations can be deleted for a clear viewing pane. The main two components of the interface consist of:

- Mentee Side (Tablet): This is treated as an end user interface, and no image processing computations occur on this device. The tablet is the key tool to show and fetch the image at the front end as well as a communication interface between mentee and mentor. It also operates as the server for connection purposes.
- Mentor Side (PC): This is where the main software for processing the detection, annotation, and posting to the tablet resides. The software interface has the following functions: crop a template, create an annotation, track the template, and send the calculated annotation positions to the tablet.

The challenge of running at near-real-time is solved by a three thread parallel computing architecture. The first thread serves to pull in the video frames from the tablet so that the mentor has a clean feed as well as to facilitate inter-thread communication. The second thread handles the bulk of the calculations. It is in this thread that the processing algorithms work to detect, match, and translate points. The third thread is responsible for the communication with the tablet to ensure the most up to date annotations are displayed.

As the majority of the computational load exists in the second thread, this paper focuses on the algorithms of this thread. When the mentor selects a template, the system automatically detects the features in the template image. The locations of those template features are saved as T along with the annotation points (A) made on the template image. Then, for each iteration of the computational thread, a frame has its feature points likewise detected and stored in S - a



(a) The mentor annotating points to be displayed (b) The tablet displaying the annotated field of view (c) The mentee looking through the tablet at the annotated surgical site

Fig. 2: The developed system

Algorithm 1 Template and scene keypoint matching

Annotation points: $A = \{(x_{ak}, y_{ak})\}, k \in [1, v]$
Template feature points: $T = \{(x_{ti}, y_{ti}, f_{ti})\}, i \in [1, m]$
Scene feature points: $S = \{(x_{sj}, y_{sj}, f_{sj})\}, j \in [1, n]$
for $i \in [1, m]$ **do**
  for $j \in [2, n]$ **do**
    if $f_{ti} \cong f_{sj}$ **then**
      $q \leftarrow (i, j)$
    end if
  end for
  if q exists **then**
    $M \leftarrow q$
  end if
end for

Algorithm 2 Extracting parameters for projection

Top-left crop point for template: (x_c, y_c)
for $(i, j) \in M$ **do**
  for $(i, j) \in M$ where $index(\hat{i}, \hat{j}) > index(i, j)$ **do**
    $D_T \leftarrow \sqrt{(x_{ti} - x_{t\hat{i}})^2 + (y_{ti} - y_{t\hat{i}})^2}$
    $D_S \leftarrow \sqrt{(x_{sj} - x_{s\hat{j}})^2 + (y_{sj} - y_{s\hat{j}})^2}$
    $\theta \leftarrow \tan^{-1}(\frac{x_{ti}-x_{t\hat{i}}}{y_{ti}-y_{t\hat{i}}}) - \tan^{-1}(\frac{x_{sj}-x_{s\hat{j}}}{y_{sj}-y_{s\hat{j}}})$
  end for
end for
$\bar{d}_t \leftarrow median(D_T)$; $\bar{d}_s \leftarrow median(D_S)$
$r \leftarrow \frac{\bar{d}_s}{\bar{d}_t}$
$a \leftarrow median(\theta)$
$\bar{x}_t \leftarrow mean(x_t)$; $\bar{y}_t \leftarrow mean(y_t)$; $\bar{x}_s \leftarrow mean(x_s)$; $\bar{y}_s \leftarrow mean(y_s)$
Project points $\in A$ using Equation (1)

second keypoint array. Algorithm 1 shows how each of the sets are compared to find matching sets between the two keypoint arrays. This algorithm results in an array M of matching indexes. Using the set of matches M, along with T and S, Algorithm 2 finds the changes in pan shift, rotation, and scale. For each cloud of matched keypoints, the distances between every point pair (D_T and D_S) and the difference in angles between each corresponding point pair across (θ) is determined. The ratio (r) of sizes between the template and current scene comes from the median distances in D_T and D_S. The system then finds the centroids of each of the matched points clouds. All these values are used to find the projection locations of the annotations (P) by applying Equation (1) to each of k annotation points.

$$P_k = \left(\hat{x}_s - \frac{\cos(\alpha)(-x_{ak} + x_c + \hat{x}_t)}{r} + \frac{\sin(\alpha)(-y_{ak} + y_c + \hat{y}_t)}{r}\right), \left(\hat{y}_s - \frac{\sin(\alpha)(-x_{ak} + x_c + \hat{x}_t)}{r} + \frac{\cos(\alpha)(-y_{ak} + y_c + \hat{y}_t)}{r}\right) \tag{1}$$

Feature detection was chosen over tracking due to the massive amounts of occlusion surrounding the key features in a surgical context. Frame-by-frame trackers such as Lucas-Kanade lose or misinterpret tracking points too quickly to be useful in this system. As another disregarded option, template matching constrains the detections to replicas of the template image. However, continuous feature detection allows for template matching without perfect information and scene changes, and is robust during and after occlusion. This makes it an optimal choice for our surgical context. There are two main feature detection algorithms: SURF and SIFT. While the algorithmic differences between the two are outside the scope of this paper, the main difference is that SURF is faster while SIFT detects more features [24]. The evaluation presented below was conducted to test the performance of these two algorithms under the particular use conditions and environment of the application's context.

On the client side where the server runs, the client receives a sequence of data through an *http* form for communication. For every post action, values are received from the tablet to instruct how the annotation string must be decoded. From this, the sequence of points is extracted, and the desired overlay information is re-rendered on the current view at the mentee side generating the augmented reality.

B. Evaluation

The crucial aspect of this system relies on tracking precision and annotation frame rate. To assess performance based on these two criteria, two state of the art feature detection algorithms were chosen to perform the tracking: Scale Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF). These each take a parameter known as a hessian value that determines how descriptive a given point is. The stronger the point, the greater the hessian; therefore as the threshold goes up, the number of detected points goes down while their

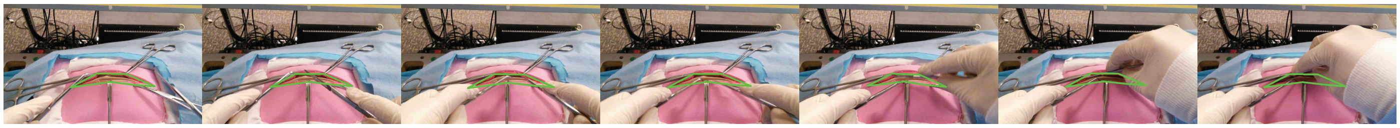


Fig. 3: Two example video strips (three frames apart each) taken for the evaluation

TABLE I: The list of control variables

Variables	Parameters	Values
X_1	Feature Detector	SURF / SIFT
X_2	Matcher	Brute force / FLANN
X_3	Hessian Threshold	0 - 2500 (100 step increments)
X_4	Video Contexts	Stationary, Pan, Zoom, Skew, Minor occlusion, Major occlusion

strength goes up. In order to perform complete tracking, feature matchers were used to find good matches between the template and the target image view. Thus, a brute force matcher and Fast Library for Approximate Nearest Neighbors (FLANN) matcher were used to determine which settings result in the best performance. The experimental procedure is based on evaluating the combinations of different feature detectors and matchers (see Table I) in different video contexts.

The system's annotation update rate was determined by C/T with $C = 50$ calculated frames and T being measured as the time taken to post those 50 frames. This reflects how often mentees has their annotations updated with the most current information. Each 50 frames constituted 1 trial, and 25 trials were collected for each of the combinations of the system parameters: feature detector type, matcher, and hessian threshold (X_1, X_2, and X_3 shown in Table I). The video was held stationary for each of these trials.

The other performance measure studied was tracking accuracy. This was tested with all four tracking parameter combinations, each with the wide range of hessian values (Table I). To test the accuracy, three video sequences (a total of 456 frames) were saved and manually annotated using LabelMe [25]. The videos simulated different contextual uses (the X_4 parameter in Table I). These videos were collected from the tablet as a robotic arm held and manipulated its positions in a controlled and pre-programmed fashion. The first video incorporated slow movements (20 *mm/s*) by the robotic arm and conducted panning, zooming, and skewing motions. The second mirrored the first but ran at 50 *mm/s*. The third and final video showed a stationary video with minor occlusion (surgical tools), major occlusion (tools and hands), and no occlusion at all. These image sequences (without annotations) were then fed into the system in lieu of the tablet video stream. Fig. 3 shows such a filmstrip incorporating occlusion. In these filmstrips, the annotated points were the two edges of a simulated incision and the four surgical tools holding that incision open. For each frame, the differences in the posted annotation values and the corresponding a priori hand-annotated values were squared and averaged to find the Mean Squared Error (MSE) result.

IV. Results

A. Update Rate

The update rate plot of all four algorithms (Fig. 4) is presented as a function of the hessian value of the detector.

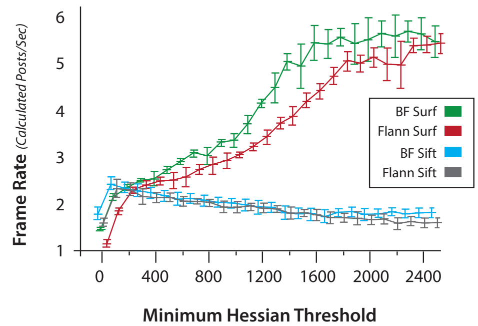


Fig. 4: Experiment 1 - Update rates against different algorithms

All four algorithms have similar slopes until the curves change at a 200-300 hessian value. The curves for SIFT decline after reaching that point, while the SURF detectors continue increasing albeit at a slower pace. In addition, it is notable that as the hessian threshold value increases, the SURF detectors' variability increases as well. Interestingly, the SIFT detector's variability remained relatively low compared to the SURF algorithms. Finally, the SURF detectors reach up to around 6 updated frames/sec in sharp contrast the SIFT detectors, which reach a peak around 2.5 updated frames/sec.

B. Tracking Accuracy

While the data for update rates shows some clear trends, the accuracy data is far noisier. The following three graphs (Fig. 5a, Fig. 5b and Fig. 5c) show the recognition accuracy for each sequence clip according to the best average overall algorithm: a SURF detector at 1500 hessian with a FLANN matcher. The graphs bin 5 frames together to show the average performance trends throughout the video as different tasks were performed. When accuracy was compared between matchers and against hessian thresholds, it was found that SURF and SIFT have means on the same order of magnitude. However, upon further inspection, it was discovered that the high average MSE for the SURF detectors comes from spikes on the frames of incredibly large error lasting a single frame at a time. Anderson-Darling normality tests run on the data found the SURF detectors to be non-normal (p-value > 0.05) while the SIFT detectors were found to be normal (p-value $<= 0.05$). Fig. 6 shows these differences from the means and medians, along with the different standard deviations for the sets.

V. Discussion

In order to find the most adequate algorithm for the telementoring system, the rates of four algorithms have been tested. As shown in Fig. 4, brute force SURF was the fastest among the tested algorithms. The hessian value threshold indirectly influenced the number of points to show on the frame by filtering out poor features. According to this, it is

reasonable that matching fewer points was faster than many points. However, the fact that SIFT does not have nearly the increase in update rate seems to confound this logic. In any case, the speed of each algorithm is only important when the algorithm is able to adequately track the template.

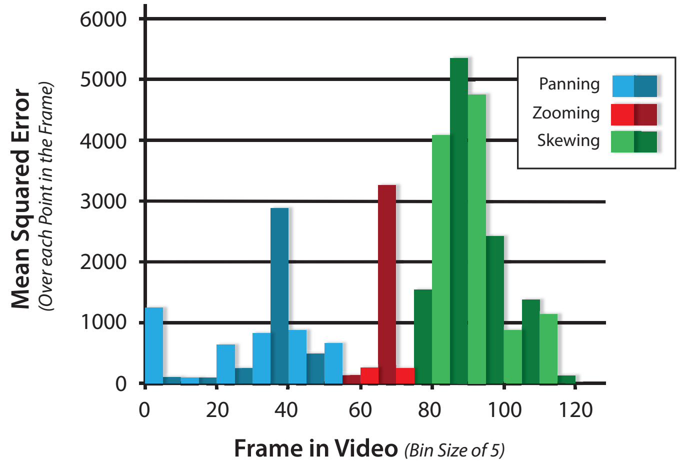
Fortunately, SURF excelled in tracking accuracy beyond SIFT as well. Although less stable with large single-frame errors, the SURF detector-based system corrected itself quickly and showed a much lower median than SIFT. A trade-off between speed and accuracy was expected; however, none of the algorithms showed a statistically strong correlation (p-values all > 0.05). It should be noted that for video 3, when the view is stable and only occlusion is applied, the tracking is very accurate (MSE of 25.34 pixels^2 for minor occlusion and 94.87 pixels^2 for major occlusion). It is assumed that the main use scenario would align greater with this video context than the first 2 videos; if the robotic arm was moving, for the surgeon to use the system they would have to be moving with the tablet.

The next step of evaluation is to include human subjects as part of the contextual testing. This will be done with the best parameter combination found in the current study, and usability metrics will be evaluated. In the future, such a study with surgeons will shed light on whether such update rates and accuracies are acceptable for the task. A small limitation comes from the small sample size in experiment 2. In addition, the results would benefit from a wider sampling of real or simulated surgical scenarios (longer videos, different surgical regions and tools, etc.). In terms of stability, the system presented can detect and compensate for movement (skew, and in plane rotation) within a range, however, it works best when stable. Therefore holding the tablet with human hands may have some impact on the tracking accuracy, because humans cannot hold a video perfectly still. This scenario should be tested to assess the extent of the impact of a human holding the tablet. Finally, surgical environments are meant to be sterile. While sterilizing a tablet computer is problematic, placing it in a clear plastic bag may allow acceptable levels of sterility. It is currently unknown to what degree this solution or other similar solutions would impact the integrity of the system's design

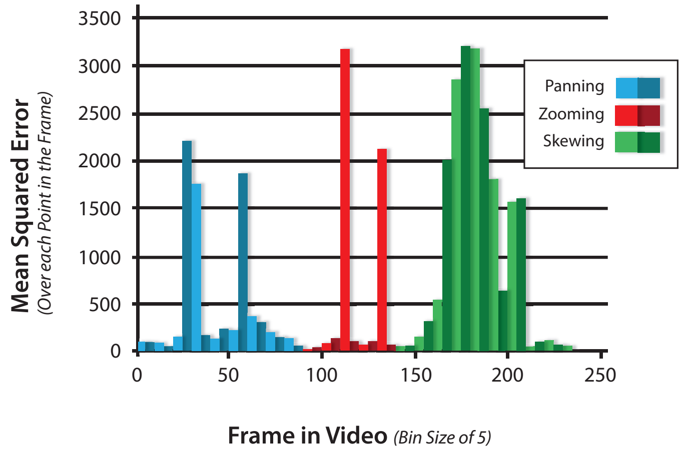
This work serves as the base of the system's design, but as time progresses, features to be implemented include body tracking and gestural interaction with the robotic arm to seamlessly integrate the tablet into the environment, the addition of surgical tool image overlays and other annotations beyond point-sketching, and work on the mentor interface to increase the mentor's sense of telepresence. With these additions, the system should become more contextually generalizable. Working with surgeons and training hospitals will help ensure that the features to be added will indeed achieve these goals.

VI. Conclusion

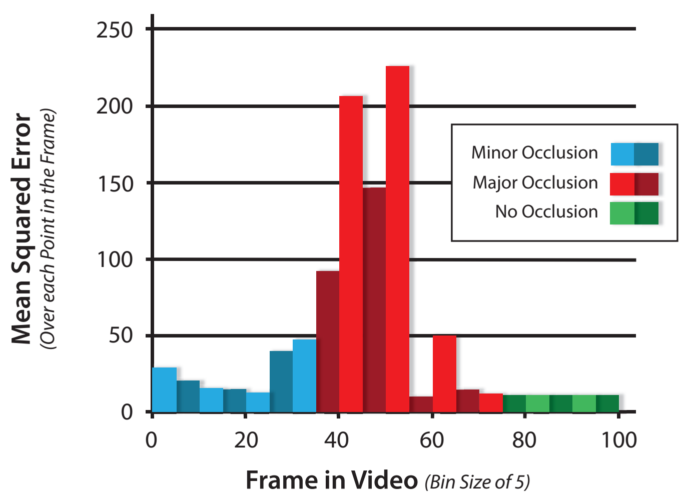
There are many contexts in which surgical telementoring and augmented reality can come together to provide value to patients and physicians alike. The development of such a system is challenging, yet not impossible. In this work, a prototype system was developed and presented, and data



(a) Video 1: Fast Movement - Each bar represents the average over 5 frames



(b) Video 2: Slow Movement - Each bar represents the average over 5 frames



(c) Video 3: Static with Occlusion - Each bar represents the average over 5 frames

Fig. 5: Experiment 2 - Tracking error over various video contexts for an optimal tracker

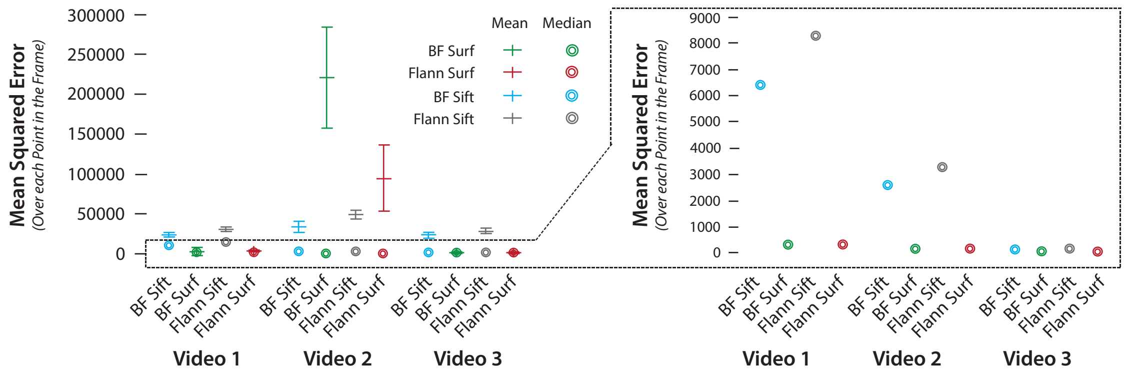


Fig. 6: Experiment 2 - Comparisons between algorithms' mean squared errors (MSE) over all hessian thresholds parameterized by mean and median

on the various parameters that went into the system design were collected. It was found that the tracking module when implemented with SURF was superior to SIFT in speed and accuracy, with an optimal hessian threshold at 1500. Within SURF, it seems that FLANN is slightly more accurate while being slightly slower. While this is contrary to our original ideas, it provides insights into the nature of the matchers in this contexts, and justifies our decisions to test these differing parameters systematically. Going forward, the designed system will continue to be improved and tested with users in surgical contexts of training and consulting.

ACKNOWLEDGMENTS

We would like to acknowledge the open source projects critical to our work: OpenCV, Android-Eye, cURL, and LabelMe.

REFERENCES

[1] A. Q. Ereso *et al.*, "Live transference of surgical subspecialty skills using telerobotic proctoring to remote general surgeons," *Journal of the American College of Surgeons*, vol. 211, no. 3, pp. 400–411, Sep 2010.

[2] D. Wood, "No surgeon should operate alone: How telementoring could change operations," *Telemedicine and e-Health*, vol. 17, no. 3, pp. 150–152, Apr 2011.

[3] G. H. Ballantyne, "Robotic surgery, telerobotic surgery, telepresence, and telementoring," *Surg Endosc*, vol. 16, no. 10, pp. 1389–1402, Oct 2002.

[4] B. Challacombe, L. Kavoussi, A. Patriciu, D. Stoianovici, and P. Dasgupta, "Technology insight: telementoring and telesurgery in urology," *Nat Clin Pract Urol*, vol. 3, no. 11, pp. 611–617, Nov 2006.

[5] B. R. Lee and R. Moore, "International telementoring: a feasible method of instruction," *World journal of urology*, vol. 18, no. 4, pp. 296–298, 2000.

[6] L. H. Eadie, A. M. Seifalian, and B. R. Davidson, "Telemedicine in surgery," *British Journal of Surgery*, vol. 90, no. 6, pp. 647–658, 2003.

[7] M. B. Shenai *et al.*, "Virtual interactive presence and augmented reality (VIPAR) for remote surgical assistance." *Neurosurgery*, 2011.

[8] J. B. Harris and G. Jones, "A descriptive study of telementoring among students, subject matter experts, and teachers: Message flow and function patterns," *Journal of research on computing in education*, vol. 32, pp. 36–53, 1999.

[9] M. A. Price and H.-H. Chen, "Promises and challenges: Exploring a collaborative telementoring programme in a preservice teacher education programme," *Mentoring and Tutoring*, vol. 11, no. 1, pp. 105–117, 2003.

[10] C. A. Kasprisin, P. B. Single, R. M. Single, and C. B. Muller, "Building a better bridge: Testing e-training to improve e-mentoring programmes in higher education," *Mentoring and Tutoring*, vol. 11, no. 1, pp. 67–78, 2003.

[11] B. J. Dixon *et al.*, "Surgeons blinded by enhanced navigation: the effect of augmented reality on attention," *Surgical endoscopy*, vol. 27, no. 2, pp. 454–461, 2013.

[12] K. M. Augestad *et al.*, "Clinical and educational benefits of surgical telementoring," in *Simulation Training in Laparoscopy and Robotic Surgery*. Springer, Jan 2012, pp. 75–89.

[13] R. Azuma *et al.*, "Recent advances in augmented reality," *IEEE Computer Graphics and Applications*, vol. 21, no. 6, pp. 34–47, Nov 2001.

[14] R. K. Miyake *et al.*, "Vein imaging: a new method of near infrared imaging, where a processed image is projected onto the skin for the enhancement of vein treatment," *Dermatologic surgery*, vol. 32, no. 8, pp. 1031–1038, 2006.

[15] A. N. Healey, C. P. Primus, and M. Koutantji, "Quantifying distraction and interruption in urological surgery," *Qual Saf Health Care*, vol. 16, no. 2, pp. 135–139, Apr 2007.

[16] "FraunhoferMEVIS: Mobile Liver Explorer." [Online]. Available: http://www.mevis.fraunhofer.de/en/solutions/mobile-liver-explorer.html

[17] H. Fuchs *et al.*, "Augmented reality visualization for laparoscopic surgery," in *Medical Image Computing and Computer-Assisted Intervention*. Springer, 1998, pp. 934–943.

[18] P. L. Kubben, "Neurosurgical apps for iPhone, iPod touch, iPad and android," *Surgical neurology international*, vol. 1, no. 1, p. 89, 2010.

[19] T. Eguchi *et al.*, "Three-dimensional imaging navigation during a lung segmentectomy using an iPad," *European Journal of Cardio-Thoracic Surgery*, vol. 41, no. 4, pp. 893–897, 2012.

[20] J. J. Rassweiler *et al.*, "iPad-assisted percutaneous access to the kidney using marker-based navigation: initial clinical experience," *European urology*, vol. 61, no. 3, pp. 628–631, 2012.

[21] B. G. Lindeque, O. I. Franko, and S. Bhola, "iPad apps for orthopedic surgeons," *Orthopedics*, vol. 34, no. 12, pp. 978–981, 2011.

[22] R. J. Rohrich, D. Sullivan, E. Tynan, and K. Abramson, "Introducing the new PRS iPad app: The new world of plastic surgery education," *Plastic and reconstructive surgery*, vol. 128, no. 3, pp. 799–802, 2011.

[23] F. Volont, J. H. Robert, O. Ratib, and F. Triponez, "A lung segmentectomy performed with 3D reconstruction images available on the operating table with an iPad," *Interactive cardiovascular and thoracic surgery*, vol. 12, no. 6, pp. 1066–1068, 2011.

[24] L. Juan and O. Gwun, "A comparison of sift, pca-sift and surf," *International Journal of Image Processing (IJIP)*, vol. 3, no. 4, pp. 143–152, 2009.

[25] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "LabelMe: a database and web-based tool for image annotation," *Int J Comput Vis*, vol. 77, no. 1-3, pp. 157–173, May 2008.

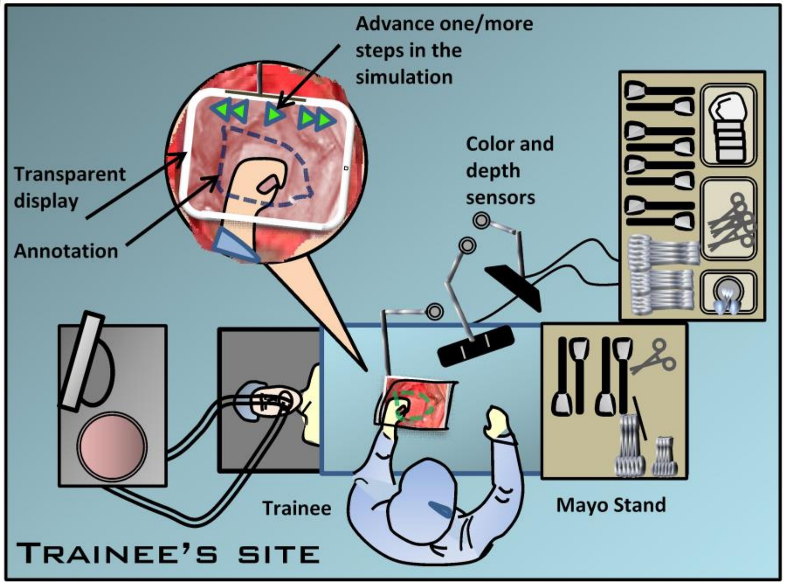
See-What-I-Do: Increasing mentor and trainee sense of co-presence in trauma surgeries with the STAR platform

FY13 Medical Practice Initiative (MPI) Augmented Reality for Medical Applications (ARM)

DMRDP

PI: Juan P Wachs **Org:** Purdue University

- **Problem**: Trauma treatment requires the immediate integration of expertise and experience of multiple specialists. Telementoring systems are promising but presently limited.
- **Hypothesis**: Increasing the mentor's and trainee's sense of co-presence during telementoring using AR increases objective and subjective measures of the trainee's surgical performance.
- **Military Relevance**: The proposed system improves care at forward-based medical facilities by bridging the experience and expertise gap for recent medical training graduates and for injuries requiring multiple surgical expertise.



Proposed Solution

- An **system for telementoring based on AR** (STAR) with a patient-size gesture-based platform at the mentor site and with a platform providing actionable visual information of current and next steps of the procedure at the trainee site.
- Real-time depth+color data is acquired at the **trainee site**.
- Data is annotated graphically through a gesture recognition interface at the **mentor site**. Data also seeds a simulation of the procedure.
- At the **trainee site**, annotations and simulation visualization augment the trainee's view of the actual surgical field seamlessly using a transparent display, illustrating the current *and next* steps of the procedure.

Timeline and Total Cost (direct and indirect)

Activities FY	1	2	3	4	5
Task 1: R&D a patient-size interaction platform with a UI based on gesture recognition					
Task 2: R&D a transparent display AR system illustrating current & next steps of procedure					
Task 3: Validate & refine STAR in controlled env. for cricothyroidotomy on human simulator					
Task 4: Validate refined STAR in a simulated Role 2 FOB (austere env.) for damage control laparotomy					
Yearly budget in $1,000's ($1,725,000 total)	254	244	248	484	493